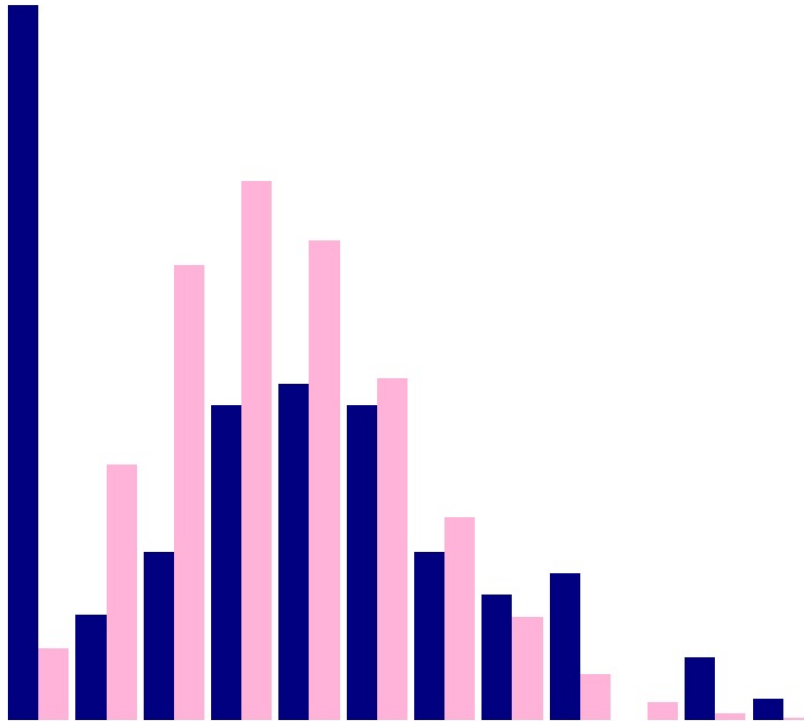


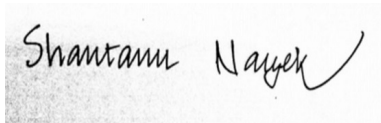
Zero- Inflated Poisson

A modification over Poisson Distribution



Shantanu Nayek
Roll no. : 452
Supervisor : Professor Surupa Chakraborty

I affirm that I have identified all my sources and that no part of my dissertation paper uses unacknowledged materials.

Student's Signature :  Date : 13 . 04 . 2022

Contents

1	Introduction	3
2	Organization of the paper	4
2.1	The Problem	4
2.2	Objectives of the paper	5
3	Collection of the data	6
4	Zero-Inflated Poisson Distribution	8
4.1	Description	8
4.2	Expectation and Variance	8
4.3	Estimation of ZIP Parameters by Method of moments	9
4.4	Estimation of ZIP Parameters by Method of maximum likelihood . .	10
5	Fitting the data	11
5.1	Fitting of Poisson Distribution	11
5.2	Fitting of ZIP Distribution	12
5.3	Justification of preferring ZIP Distribution	13
5.3.1	Graphical representation using Column Diagram	13
5.3.2	Test of the parameter π	14
5.3.3	Comparision of Standard Errors	15
6	Estimation of Standard Errors of the Estimators of ZIP parameters	16
6.1	Estimation of Standard Errors of Maximum likelihood estimators of π and λ using Fisher Information Matrix	16
6.2	Estimation of Standard Errors of Method of moment estimators of π and λ using Delta Method	18
6.3	Tabular representation of outputs	20
6.4	Observation	21
7	Conclusion	21
8	Acknowledgement	22
9	Appendix	23

1 Introduction



Given a count data in hand and if the count seems to be unbounded, statisticians are more inclined towards fitting a Poisson Model blindfold in most cases. But this may sometimes lead to erroneous conclusions regarding analysis or other statistical works based on the data. We know that for a Poisson Model, the mean and variance for a sample are very close in general, but for the data it may not be. There may be another problem like if there is a value (say 0 or 1 or 2 or any integer) which has a very high frequency in the sample, then an inflation may be present at the probability curve at that particular value. Based on the above situation, in this study, data on number of seminars/webinars attended by college students during a 2019-2021 was collected. Since it was an unbounded count data, initially we showed interest to fit a Poisson Model for analysis of the data. But it was observed some excess zeroes were there. This presence of excess zeroes, led to thinking of modification of the Poisson Model and we opt to work with the Zero Inflated Poisson distribution. Various measures were taken for justification of fitting a ZIP over a Poisson Model. Some of the justifications are, testing for the parameter π , standard errors for the two models, etc. On giving preference to the ZIP model, I intend to estimate the parameters of the model by Method of Maximum Likelihood and Method of Moments. Finally, there were interests in comparing the efficacy of the estimation procedures.

2 Organization of the paper



2.1 The Problem

Firstly, it is our natural tendency to opt for a Poisson Model for dealing with count data. But count data does not always satisfy a Poisson Model. In case, if we do such assumptions and do further analysis we will lead to some outcomes and observations. But those may not be appropriate. As a result of which the conclusions obtained will be vague and misleading in many situations. So, to increase the consciousness among the statisticians regarding choosing a model for such unbounded count data is a point of interest.

Secondly, in this paper I worked with a real life data on no. of webinars/seminars attended by college students in the span of 2019-2021. In this data fitting a Zero-Inflated Poisson Model over Poisson Model gives a better conclusions rather more appropriate conclusions. Some justifications were shown regarding the selected model over the another.

Thirdly, the parameters of the Zero Inflated Poisson Distribution were estimated by two estimation procedures namely, Method of moments and Method of maximum likelihood. Then a study for determining the better estimation procedure is done in terms of their standard errors.

2.2 Objectives of the paper



The paper has two major objectives. To increase consciousness in choosing a model for a count data correctly. The statisticians are blindfold in fitting a Poisson Model getting a count data in hand , but in practice it is not always perfect. Taking wrong assumptions of statistical model may lead to erroneous conclusions of statistical analyses. As a consequence of such difficulty, an illustrative real life sample is collected and with certain justifications it is shown that fitting of a Poisson Model is not a good job inspite of the data being an unbounded count data. For that a slight modification is done over the Poisson Distribution and fitted again. The modified distribution is Zero -Inflated Poisson Distribution. And we want to study whether there is any improvement while fitting this new model.

There are two parameters of Zero-Inflated Poisson Distribution. They are estimated by two well-known estimation methods , Method of Moments and Method of Maximum Likelihood. The two estimation procedures are compared in terms of their standard errors to conclude which procedure is more efficient for estimation purpose. There will be some difficulties in computing the standard errors, making some assumptions we proceed to derive approximate expressions and finally the exact value of the standard errors for the collected sample.

3 Collection of the data



Motive : In order to handle the misconception that statisticians have regarding fitting of count data , this data is collected . There were certain ideas that lead to the thought that the data may have a zero inflation.

Issue : Data on number of webinars/ seminars attended by college students between 2019-2021 was collected by questionnaire method using Google Forms.

Source : Some of the major sources are :

- St.Xavier's College(Autonomous), Kolkata
- R.K.M. Residential College,Narendrapur
- Haldia Government College
- Techno India University,Kolkata
- St.Stephens College,Delhi
- Calcutta Medical College
- Jadavpur University
- M.U.C. Women's College, Burdwan

Time of Collection : The data was collected in between 28/09/2021. -30/09/2021

Description of the data : No. of webinars/seminars attended by college students during the span of 2019-2021 was collected. The data also consists of the Gender, Academic Degree, Stream of Study, Current Year (as of September 2021).In this paper we are just interested in the count of no. webinars/seminars attended. The other information are collected for future studies(if essential).

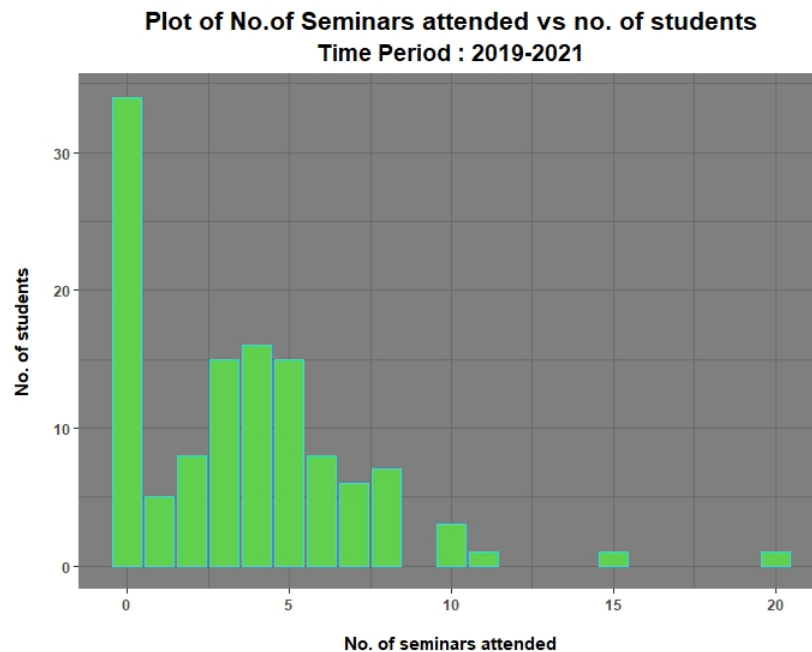
No. of respondents : 120

Factors leading to Zero- Inflation : There were a number of factors that led to the thought that the data may be zero inflated -

- There may be certain colleges where seminars or webinars are not organised for the students that may be due to lack of infrastructure or other reasons.
- Students might not show intention to attend the seminars as he/she might feel to have well preparation on that topic.
- In many cases, students are not interested in filling long forms for attending seminars.

N.B. There may be some other reasons behind zero-inflation. The above mentioned are some of the significant reasons.

Visualisation of the Sample data :



4 Zero-Inflated Poisson Distribution

4.1 Description

Diane Lambert's Zero-inflated Poisson model concerns a random event containing excess zero-count data in unit time. The ZIP distribution with parameters π and λ , is denoted by ZIP (π, λ) .

The pmf of zero-inflated Poisson Distribution is given by-

$$P(X = x) = \begin{cases} \pi + (1 - \pi)e^{-\lambda}, & \text{if } x = 0 \\ \frac{(1-\pi)e^{-\lambda}\lambda^x}{x!}, & \text{if } x = 1, 2, 3, 4, \dots \end{cases}$$

The parameter π gives the extra probability thrust at the value 0; when it vanishes, ZIP (π, λ) reduces to Poisson(λ). The zero-inflated Poisson (ZIP) model mixes two zero generating processes. The first process generates zeros. The second process is governed by a Poisson distribution that generates counts, some of which may be zero.

4.2 Expectation and Variance

$$\begin{aligned} E(X) &= \sum_{x=0}^{\infty} [x.P(X = x)] \\ &= 0.[\pi + (1 - \pi).e^{-\lambda}] + (1 - \pi). \sum_{x=1}^{\infty} \left[\frac{x.e^{-\lambda}\lambda^x}{x!} \right] \\ &= (1 - \pi)\lambda \end{aligned}$$

$$Var(X) = E(X^2) - E(X)^2$$

$$\begin{aligned} \text{Now, } E(X^2) &= \sum_{x=0}^{\infty} [x^2 \cdot P(X = x)] \\ &= 0 \cdot [\pi + (1 - \pi) \cdot e^{-\lambda}] + (1 - \pi) \cdot \sum_{x=1}^{\infty} \left[\frac{x^2 \cdot e^{-\lambda} \lambda^x}{x!} \right] \\ &= (1 - \pi)(\lambda + \lambda^2) \end{aligned}$$

$$Var(X) = \lambda(1 - \pi)(1 + \lambda\pi)$$

4.3 Estimation of ZIP Parameters by Method of moments

$$\text{Define , } \bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}$$

$$\text{Clearly, } (1 - \pi)\lambda = \bar{X}, \quad \lambda(1 - \pi)(1 + \lambda\pi) = S^2 \quad \dots\dots(*)$$

On solving (*), we obtain

$$\widehat{\lambda_{MME}} = \bar{X} + \frac{S^2}{\bar{X}} - 1 \quad \widehat{\pi_{MME}} = \frac{S^2 - \bar{X}}{\bar{X}^2 + S^2 - \bar{X}}$$

Now, the estimate of \bar{X} can be greater than S^2 . So, $\widehat{\pi_{MME}}$ may take negative values. But, $0 < \pi < 1$.

A slight modification is done over the obtained estimators. The resultant corrected Method of Moments Estimators are :

$$\hat{\lambda} = \begin{cases} \bar{X}, & \text{if } \bar{X} \geq S^2 \\ \widehat{\lambda_{MME}}, & \text{if } \bar{X} \leq S^2 \end{cases}$$

$$\hat{\pi} = \begin{cases} 0, & \text{if } \bar{X} \geq S^2 \\ \widehat{\pi_{MME}}, & \text{if } \bar{X} \leq S^2 \end{cases}$$

4.4 Estimation of ZIP Parameters by Method of maximum likelihood

The likelihood function is given by -

$$L(\pi, \lambda) = \prod_{i=1}^n P(X_i = x_i)$$

Let, Y be the number of X_i 's taking value 0. The likelihood function can be re-written as :

$$L(\pi, \lambda) = (\pi + (1 - \pi)e^{-\lambda})^Y \prod_{i=1(1)n, X_i \neq 0} \frac{(1 - \pi)e^{-\lambda} \lambda^{x_i}}{x_i!}$$

The log likelihood function is given by :

$$\begin{aligned} \log(L(\pi, \lambda)) &= Y \log(\pi + (1 - \pi)e^{-\lambda}) + (n - Y) \log(1 - \pi) - (n - Y)\lambda \\ &+ n\bar{X} \log(\lambda) - \log\left(\prod_{i=1}^n X_i\right) \dots\dots\dots(**) \end{aligned}$$

Differentiating (**) with respect to λ and equating it to 0 , we obtain :

$$\frac{\delta \log(L(\lambda, \pi))}{\delta \lambda} = 0$$

$$\text{or, } \frac{n\bar{X}}{\lambda} = (n - Y) + \frac{Y(1 - \pi)e^{-\lambda}}{(\pi + (1 - \pi)e^{-\lambda})} \dots\dots\dots(\text{i})$$

Differentiating (**) with respect to π and equating it to 0 , we obtain :

$$\frac{\delta \log(L(\lambda, \pi))}{\delta \pi} = 0$$

$$\text{or, } \frac{Y(1 - \pi)(1 - e^{-\lambda})}{(\pi + (1 - \pi)e^{-\lambda})} = n - Y \dots\dots\dots(\text{ii})$$

On solving (i) and (ii) , the Maximum likelihood estimators , say $\widehat{\pi}_{MLE}$ and $\widehat{\lambda}_{MLE}$ are obtained by iterative method.

Note :

For our collected sample (n=120), the estimates are tabulated below :

Table 1: Estimates of π and λ by MME and MLE

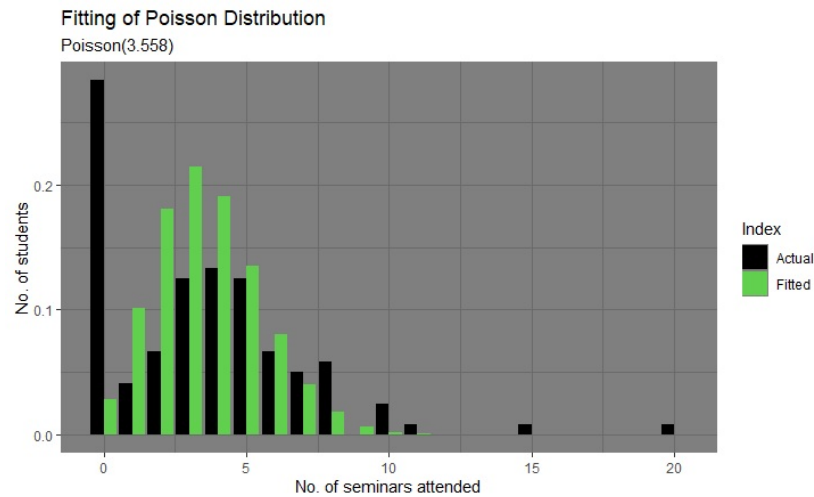
Parameter	Method	Estimate
π	MME	0.3832
π	MLE	0.2781
λ	MME	7.7691
λ	MLE	4.9292

5 Fitting the data

5.1 Fitting of Poisson Distribution

Since, it is an unbounded count data ,we showed our initial interest to fit Poisson distribution.

From the sample , we obtained the method of moments estimate of the parameter λ for fitting $\text{Poisson}(\hat{\lambda})$. the obtained value of $\hat{\lambda}$ is 3.558.

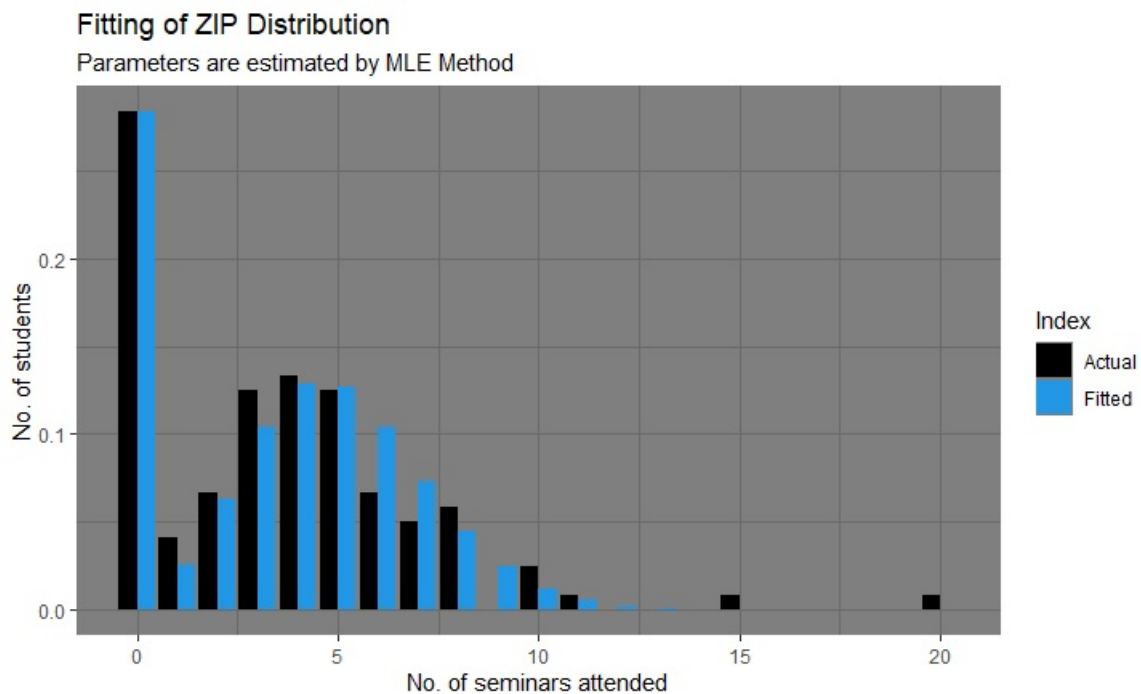


Observation : The fitted probability and the relative frequency at zero point deviates the most. The fit underestimates the probability at 0 and overestimates almost rest of the probabilities. So, we wanted to be concerned a bit about that zero point.

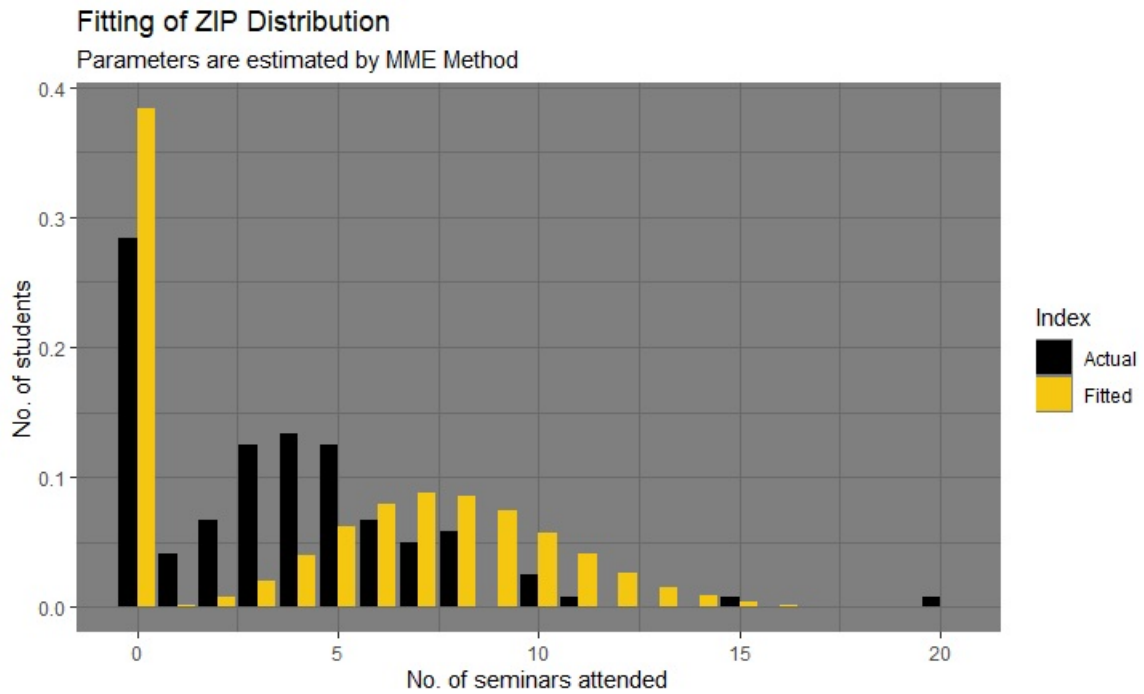
5.2 Fitting of ZIP Distribution

Since, the point zero was underestimated in the above fit, so we showed interest to fit Zero-Inflated Poisson Distribution. Zero Inflated Poisson Distribution is obtained by slight modification over Poisson Distribution , just providing some excess probability thrust to the point zero.

For fitting a ZIP distribution, the parameters are taken as the estimates obtained by Method of Maximum Likelihood (From Table 1).

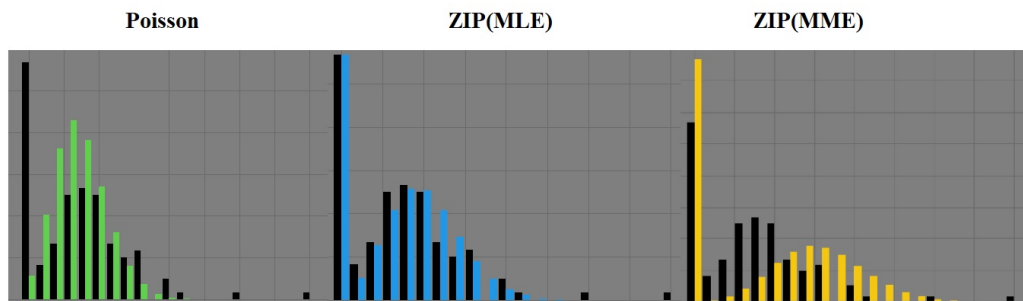


For fitting a ZIP distribution, the parameters are taken as the estimates obtained by Method of Moments(From Table 1).



5.3 Justification of preferring ZIP Distribution

5.3.1 Graphical representation using Column Diagram



Observations :

- i) Fitting of ordinary Poisson Distribution underestimates probability at zero point and overestimates the probability in almost all other points.
- ii) Fitting of ZIP Distribution , where parameters are estimated by Method of Moments overestimates the probability at zero and underestimates the probabilities in almost all other points.
- iii) Apparently, fitting of ZIP Distribution, where the parameters are estimated by Method of Maximum Likelihood gives better results than the previous two cases.

5.3.2 Test of the parameter π

Let , π be the parameter of the Zero - Inflated Poisson Distribution which gives excess probability thrust to the zero point.

To test $H_0 : \pi=0$ **against** $H_1 : \pi \neq 0$

Sample : The collected sample from the $n(= 120)$ individuals is the chosen sample for the testing procedure.

Assumption : $\sqrt{n}(\hat{\pi} - \pi) \rightarrow A.N.(0, \sigma^2)$ A.N. : Asymptotic Normal

Test Statistic : $Z = \frac{\sqrt{n}(\hat{\pi} - \pi)}{\sigma^*}$

Under H_0 , $Z = \frac{\sqrt{n}(\hat{\pi})}{\sigma_{*o}}$ where,

σ_{*o} is unknown. So, we take the estimate of σ_{*o} as the standard error of $\hat{\pi}$

obtained by Method of Maximum Likelihood.

Note :

- The estimate of $\hat{\pi}$ is obtained by Method of Maximum Likelihood.
- The Standard Error computation is explained in Section 6.1 in details with the help of Fisher Information Matrix.

Computation : For the sample ,

$$n=120, \quad \hat{\pi} = 0.2781128, \quad \sigma_{*o} = 0.04145$$

$$Z(\text{observed})=6.71497, \quad \alpha=0.05(\text{assuming}), \quad \tau_{0.025} = 1.96$$

Decision : Since $|Z(\text{observed})| > \tau_{0.025}$, we reject H_0 at 5 percent level of significance.

Conclusion : In the light of the given data , it seems that fitting a Zero- Inflated Poisson Distribution is better than fitting a Poisson Distribution.

5.3.3 Comparison of Standard Errors

The standard error of the parameter λ under the assumption of Poisson Distribution is 0.1721. The standard error of the parameter λ under the assumption of Zero - Inflated Poisson Distribution is 0.2429.

It is to be noted that for Poisson Distribution the standard errors obtained by Method of Moments and Method of Maximum Likelihood are same. For , Zero - Inflated Poisson Distribution , we considered the standard error obtained bt Method of Maximum Likelihood. This is due to the fact that Maximum Likelihood Estimation procedure is more efficient than Moment Estimation procedure (discussed in Section 6).

Here, we observe that the standard error in case of Poisson Distribution is lesser than the standard error in case of Zero - Inflated Poisson Distribution. But in Section 5.3.1 and 5.3.2 we had extracted some evidence which gives a nod to Zero-Inflated Poisson Distribution over Poisson Distribution. But it is very interesting that standard error is higher for the Zero - Inflated Poisson Distribution. One of the major reasons of this anomaly is the number of parameters in both the distribution are not same. Hence, it is not well justified to compare standard errors in this situation.

6 Estimation of Standard Errors of the Estimators of ZIP parameters

6.1 Estimation of Standard Errors of Maximum likelihood estimators of π and λ using Fisher Information Matrix

The likelihood function is given by -

$$L(\pi, \lambda) = \prod_{i=1}^n P(X_i = x_i)$$

Let, Y be the number of X_i 's taking value 0. The likelihood function can be re-written as :

$$L(\pi, \lambda) = (\pi + (1 - \pi)e^{-\lambda})^Y \prod_{i=1(1)n, X_i \neq 0} \frac{(1 - \pi)e^{-\lambda} \lambda^{x_i}}{x_i!}$$

The log likelihood function is given by :

$$\begin{aligned} \log(L(\pi, \lambda)) &= Y \log(\pi + (1 - \pi)e^{-\lambda}) + (n - Y) \log(1 - \pi) - (n - Y)\lambda \\ &+ n\bar{X} \log(\lambda) - \log\left(\prod_{i=1}^n X_i\right) \dots\dots\dots (**) \end{aligned}$$

Differentiating (**) with respect to λ , we obtain :

$$\frac{\delta \log(L(\lambda, \pi))}{\delta \lambda} = \frac{n\bar{X}}{\lambda} - (n - Y) - \frac{Y(1 - \pi)e^{-\lambda}}{(\pi + (1 - \pi)e^{-\lambda})} \dots\dots\dots (i)$$

Again, differentiating (i) with respect to λ , we obtain :

$$\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \lambda^2} = \frac{Y\pi(1 - \pi)e^{-\lambda}}{(\pi + (1 - \pi)e^{-\lambda})^2} - \frac{n\bar{X}}{\lambda^2} \dots\dots\dots (ii)$$

Differentiating (i) with respect to π , we obtain :

$$\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \lambda \delta \pi} = \frac{Y e^{-\lambda}}{(\pi + (1 - \pi)e^{-\lambda})^2} \dots\dots\dots (iii)$$

Differentiating (**) with respect to π , we obtain :

$$\frac{\delta \log(L(\lambda, \pi))}{\delta \pi} = \frac{Y(1 - \pi)(1 - e^{-\lambda})}{(\pi + (1 - \pi)e^{-\lambda})} - \frac{(n - Y)}{(1 - \pi)} \dots\dots\dots (iv)$$

Again, differentiating (iv) with respect to π , we obtain :

$$\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \pi^2} = -\frac{Y(1-e^{-\lambda})^2}{(\pi+(1-\pi)e^{-\lambda})^2} - \frac{(n-Y)}{(1-\pi)^2} \dots\dots\dots (v)$$

We define ,

$$I_{11} = - E\left(\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \lambda^2} \right) \quad I_{22} = - E\left(\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \pi^2} \right)$$

$$I_{21} = I_{12} = - E\left(\frac{\delta^2 \log(L(\lambda, \pi))}{\delta \lambda \delta \pi} \right)$$

The Fisher Information matrix $I(\lambda, \pi)$:

$$I(\lambda, \pi) = \begin{bmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{bmatrix}, \text{ where } I_{11}, I_{12}, I_{21}, I_{22} \text{ are defined.}$$

The inverse of the Fisher Information Matrix $I(\lambda, \pi)$ is:

$$I^{-1}(\lambda, \pi) = \begin{bmatrix} I^{11} & I^{12} \\ I^{21} & I^{22} \end{bmatrix} \quad (\text{let})$$

Note : The explicit form of $I^{-1}(\lambda, \pi)$ is not shown here . Generally, for a given sample , we obtain $I(\lambda, \pi)$ explicitly. Then we obtain $I^{-1}(\lambda, \pi)$ by just taking inverse of the finite matrix $I(\lambda, \pi)$.

The standard errors are given by :

$$\text{S.E.}(\widehat{\lambda_{MLE}}) = \sqrt{I^{11}}$$

$$\text{S.E.}(\widehat{\pi_{MLE}}) = \sqrt{I^{22}}$$

6.2 Estimation of Standard Errors of Method of moment estimators of π and λ using Delta Method

The Method of Moment estimates of π and λ :

Let us define , $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$, $S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)}$

Hence , $\widehat{\lambda_{MME}} = \bar{X} + \frac{S^2}{\bar{X}} - 1$ $\widehat{\pi_{MME}} = \frac{S^2 - \bar{X}}{\bar{X}^2 + S^2 - \bar{X}}$ (*)

Note :

- The derivation of the above estimators are done in Section 4.3 .
- Since for our sample $n=120$, the sample size is quite large. Hence, we approximate $(n-1)$ in the expression of S^2 by n for simpler calculation.

Let us define ,

$$\bar{X} = \sum_{i=1}^n \frac{X_i}{n} = m'_1 \quad (\text{let}) \quad \dots\dots\dots(\text{i})$$

$$S^2 = \sum_{i=1}^n \frac{(X_i - \bar{X})^2}{(n-1)} \approx \sum_{i=1}^n \frac{X_i^2}{n} - \bar{X}^2 = m'_2 - (m'_1)^2 \quad (\text{let}) \quad \dots\dots\dots(\text{ii})$$

Combining equation (i) and (ii) with (*) , we obtain ,

$$\widehat{\pi_{MME}} = 1 - \frac{(m'_1)^2}{m'_2 - (m'_1)} \quad (\text{let}) \quad \dots\dots\dots(\text{iii})$$

$$\widehat{\lambda_{MME}} = \frac{(m'_2)}{(m'_1)} - 1 \quad (\text{let}) \quad \dots\dots\dots(\text{iv})$$

For simplicity, let

$$m'_1 = T_1 \quad m'_2 = T_2 .$$

Hence , we can write in general ,

$$\widehat{\lambda_{MME}} = f(m'_1, m'_2) \quad \widehat{\pi_{MME}} = g(m'_1, m'_2)$$

$$\text{i.e, } \widehat{\lambda_{MME}} = f(T_1, T_2) \quad \widehat{\pi_{MME}} = g(T_1, T_2)$$

We observe ,

Differentiating (iv) with respect to T1 ,

$$\frac{\delta f(T_1, T_2)}{\delta T_1} = \frac{-T_2}{T_1^2}$$

Differentiating (iii) with respect to T2 ,

$$\frac{\delta g(T_1, T_2)}{\delta T_2} = \frac{T_1^2 - 2T_1 T_2}{(T_2 - T_1)^2}$$

Differentiating (iv) with respect to T2 ,

$$\frac{\delta f(T_1, T_2)}{\delta T_2} = \frac{1}{T_1}$$

Differentiating (iii) with respect to T1 ,

$$\frac{\delta g(T_1, T_2)}{\delta T_1} = \left(\frac{T_1}{T_2 - T_1}\right)^2$$

Note:

Define , $\mu'_r = \frac{\sum_{i=1}^n x_i^r}{n}$

$$E(m'_r) = \mu'_r$$

$$\text{Var}(m'_r) = \frac{\mu'_{2r} - (\mu'_r)^2}{n}$$

$$\text{Cov}(m'_r, m'_q) = \frac{\mu'_{r+q} - (\mu'_r \mu'_q)}{n}$$

Using Delta Method ,

$$\begin{aligned}
\bullet \text{Var}(f(T_1, T_2)) &= \text{Var} (T_1) \left[\frac{\delta f(T_1, T_2)}{\delta T_1} \right]^2 + \text{Var} (T_2) \left[\frac{\delta f(T_1, T_2)}{\delta T_2} \right]^2 + \\
&\quad 2 \text{Cov} (T_1, T_2) \left[\frac{\delta f(T_1, T_2)}{\delta T_1} \frac{\delta f(T_1, T_2)}{\delta T_2} \right] \\
&= \frac{(\mu'_2 - (\mu'_1)^2)}{n} \left(\frac{-\mu'_2}{(\mu'_1)^2} \right)^2 + \frac{(\mu'_4 - (\mu'_2)^2)}{n} \left(\frac{1}{\mu'_1} \right)^2 + \frac{2(\mu'_3 - \mu'_2 \mu'_1)}{n} \left(\frac{-\mu'_2}{(\mu'_1)^3} \right) \\
\bullet \text{Var}(g(T_1, T_2)) &= \text{Var} (T_1) \left[\frac{\delta g(T_1, T_2)}{\delta T_1} \right]^2 + \text{Var} (T_2) \left[\frac{\delta g(T_1, T_2)}{\delta T_2} \right]^2 + \\
&\quad 2 \text{Cov} (T_1, T_2) \left[\frac{\delta g(T_1, T_2)}{\delta T_1} \frac{\delta g(T_1, T_2)}{\delta T_2} \right] \\
&= \frac{(\mu'_2 - (\mu'_1)^2)}{n} \left(\frac{(\mu'_1)^2 - 2\mu'_1 \mu'_2}{(\mu'_2 - \mu'_1)^2} \right)^2 + \frac{(\mu'_4 - (\mu'_2)^2)}{n} \left(\frac{\mu'_1}{\mu'_2 - \mu'_1} \right)^4 + \frac{2(\mu'_3 - \mu'_2 \mu'_1)}{n} \frac{\mu_1^2 ((\mu'_1)^2 - 2\mu'_1 \mu'_2)}{(\mu'_2 - \mu'_1)^2}
\end{aligned}$$

Hence , the estimate of the standard errors are :

$$S.E.(\widehat{\lambda_{MME}}) = \sqrt{\text{Var}(f(T_1, T_2))}$$

$$S.E.(\widehat{\pi_{MME}}) = \sqrt{\text{Var}(g(T_1, T_2))}$$

6.3 Tabular representation of outputs

Table 2: Summary of the estimators of π and λ

Parameter	Method	Estimate	S.E.	C.L.(lower)	C.L.(upper)
π	MME	0.3832	0.063	0.3832 - 1.96 * 0.063	0.3832 + 1.96 * 0.063
π	MLE	0.2781	0.041	0.2781 - 1.96 * 0.041	0.2781 + 1.96 * 0.041
λ	MME	7.7691	0.732	7.7691 - 1.96 * 0.732	7.7691 + 1.96 * 0.732
λ	MLE	4.9292	0.242	4.9292 - 1.96 * 0.242	4.9292 + 1.96 * 0.242

Note :

- S.E. : Standard Error.
- C.L. : Confidence Limits.
- The confidence limits are computed taking $\alpha = 0.05$.
- $\tau_{0.025} = 1.96$ [approximately]

6.4 Observation

- The standard error of $\widehat{\lambda_{MME}}$ is higher than of $\widehat{\lambda_{MLE}}$.
- The standard error of $\widehat{\pi_{MME}}$ is higher than of $\widehat{\pi_{MLE}}$.

7 Conclusion

From the above study , we can arrive at two major conclusions. Firstly , for a count data fitting of a Poisson Distribution may not always work good. So, getting a count data in hand we should always confirm that there is no unusual inflation at a point. There may be other factors also which may deviate from Poisson Distribution. More specifically , if there is a inflation at zero point assuming all other factors being usual, we may show our interest in fitting a Zero - Inflated Poisson Distribution over Poisson Distribution.

Secondly, considering Zero - Inflated Poisson Distribution , we computed the standard errors of the method of moments estimators and maximum likelihood estimators of the parameters. We observed that the standard errors of the Maximum Likelihood Estimators are lesser than standard errors of the Method of Moments Estimators for both situations. Moreover we observed that fitting of Zero - Inflated Poisson Distribution where estimate of the parameters are obtained by Method of Maximum

Likelihood is better than fitting of Zero - Inflated Poisson Distribution where estimate of the parameters are obtained by Method of Moments. Hence, as a whole among two estimation procedures , it seems that Method of Maximum Likelihood is more efficient than Method of Moments.

8 Acknowledgement

I would like to express my gratitude to my supervisor , respected Professor Surupa Chakraborty , who has given me this great opportunity to work on this topic and gather deeper knowledge and skills . Without her constant support and guidance the completion of this project work would not have been possible. I would also like to thank all my respected professors who have always guided and encouraged me throughout the course. I am grateful to all my friends who helped me a lot in gathering information and transforming my ideas into reality. Lastly , but not the least , I am indebted to my family for their immense love , support and guidance.

9 Appendix

Code 1: R Code for the Paper

```
1
2 #-----
3 #-----
4
5         #READING THE EXCEL SHEET
6
7 setwd("X:/")
8 getwd()
9 data=read.csv("Data Dissertation.csv");data
10
11 #-----
12 #-----
13
14         #Calling Libraries
15
16 library(ggplot2)
17 library(VGAM)
18 library(tidyr)
19
20 #-----
21 #-----
22
23         #Data Visualisation
24
25 table(data$Count)
26 table(data$Stream)
27
28 #Defining theme for plots
29 theme_new=theme(
30
31     plot.title =element_text(size=16,
32                               hjust=0.5,face="bold"),
33
34     plot.subtitle =element_text(size=14,
35                                  hjust=0.5,face="bold"),
36
37     legend.title=element_text(size=10,hjust=0.5,
```

```

38         face="bold.italic"),
39
40     axis.title=element_text(face="bold"),
41
42     axis.text=element_text(face="bold")
43
44 )
45
46 z=data$Count
47 ggplot(NULL,aes(z))+
48     geom_bar(col=5,fill=3)+theme_dark()+
49     theme_new+
50     labs(title="Plot of No.of Seminars attended
51           vs no. of students",
52          subtitle="Column Diagram",
53          x="\nNo. of seminars attended",
54          y="No. of students\n",
55          col="Index")
56
57 x=0:20
58 y=c(34,5,8,15,16,15,8,6,7,0,3,1,0,0,0,1,
59     0,0,0,0,1)
60 sum(y)
61 A=cbind(x,y);A
62
63 #-----
64 #-----
65         #METHOD OF MOMENTS ESTIMATES
66
67 count=rep(x,times=y);count
68 table(count)
69         #MEAN AND VARIANCE OF THE COLLECTED SAMPLE
70 sample_mean=mean(count);sample_mean
71
72 sample_variance=var(count);sample_variance
73
74 #-----
75 #-----
76
77         #ESTIMATES _ MME _ZIP _PIE ,LAMBDA
78

```



```

79 lambda_MME=sample_mean+(sample_variance/sample_mean)+1
80 lambda_MME
81
82 pie_MME=(sample_variance-sample_mean)/
83     (sample_mean^2+
84     (sample_variance-sample_mean));
85
86 #-----
87 #-----
88
89         #ESTIMATES _ MLE _ZIP _PIE ,LAMBDA
90
91 count
92 n=length(count)
93
94 y=length(count[count==0]);y #NUMBER OF ZEROES IN THE SAMPLE
95
96 sample_no.zero=count[count!=0];sample_no.zero #THE
97                                     #SAMPLE EXCLUDING THE ZEROES
98 x=sample_no.zero
99 count
100
101         #DEFINING THE LIKELIHOOD FUNCTION
102
103 Likelihood=function(l,p)
104 {
105
106     -sum(dzipois(count,l,p,log=T))
107 }
108
109 k=mle(Likelihood,start=list(l=lambda_MME,p=pie_MME))
110 summary(k)
111
112 #-----
113 #-----
114
115         #MLE ESTIMATES USING POISSON DISTRIBUTION
116
117 Likelihood=function(l)
118 {
119

```

```

120   -sum(dpois(count,1,log=T))
121 }
122
123 k=mle(Likelihood,start=9) #9 is a arbitrary choice
124 k
125
126 #-----
127 #-----
128
129           #FITTING OF ZIP AND POISSON DISTRIBUTION
130
131 fit_POISSON_MLE=dpois(seq(0,20,1),lambda_POISSON_MLE);fit_
    POISSON_MLE
132
133 fit_ZIP_MLE=dzipois(seq(0,20,1),lambda_MLE,pie_MLE);fit_ZIP_
    MLE
134
135 fit_ZIP_MME=dzipois(seq(0,20,1),lambda_MME,pie_MME);fit_ZIP_
    MME
136
137 y=c(34,5,8,15,16,15,8,6,7,0,3,1,0,0,0,1,0,0,0,0,1)
138 ACTUAL_prob=y/sum(y);ACTUAL_prob           #ACTUAL DATA
139
140           #Graph of Poisson fitting
141
142 datat2=data.frame(0:20,Actual=ACTUAL_prob,Fitted=fit_POISSON_
    MLE)
143 data2=pivot_longer(data2,-1,"Index",values_to="v")
144 data2
145 ggplot(data2,aes(X0.20,v,fill=Index))+
146   geom_bar(position="dodge",stat="identity")+
147   labs(title="Fitting of Poisson Distribution",
148        subtitle="Poisson(3.558)",x="No. of seminars attended",
149        y="No. of students")+theme_dark()+
150   scale_fill_manual(values=c(9,3))
151
152
153           #Graph of ZIP Fitting by MLE Method
154
155
156 data3=data.frame(0:20,Actual=ACTUAL_prob,Fitted=fit_ZIP_MLE)

```

```

157 data3=pivot_longer(data3,-1,"Index",values_to="v")
158 data3
159 ggplot(data3,aes(X0.20,v,fill=Index))+
160   geom_bar(position="dodge",stat="identity")+
161   labs(title="Fitting of ZIP Distribution",
162        subtitle="Parameters are estimated by MLE Method",
163        x="No. of seminars attended",
164        y="No. of students")+theme_dark()+
165   scale_fill_manual(values=c(9,7))
166
167         #Graph of ZIP Fitting by MME Method
168
169 data4=data.frame(0:20,Actual=ACTUAL_prob,Fitted=fit_ZIP_MME)
170 data4=pivot_longer(data4,-1,"Index",values_to="v")
171 data4
172 ggplot(data4,aes(X0.20,v,fill=Index))+
173   geom_bar(position="dodge",stat="identity")+
174   labs(title="Fitting of ZIP Distribution",
175        subtitle="Parameters are estimated by MME Method",
176        x="No. of seminars attended",
177        y="No. of students")+theme_dark()+
178   scale_fill_manual(values=c(9,7))
179
180 #-----
181 #-----
182
183         #Computation of SE from Fisher Information Matrix
184         #Let lambda=1 ,pie=p ,Y= no. of Zeroes ,n=sample size
185
186 lambda_MLE=4.9292075;lambda_MLE
187 pie_MLE=0.2781128;pie_MLE
188
189 l=lambda_MLE
190 p=pie_MLE
191
192 n=120
193 Y=34
194 Xbar=(1-p)*l
195
196 I11=((Y*p*(1-p)*exp(-l))/(p+(1-p)*exp(-l))^2)-((n*Xbar)/l^2)
197 I12=(Y*exp(-l))/(p+(1-p)*exp(-l))^2

```

```

198 I22=-((Y*(1-exp(-1))^2)/(p+(1-p)*exp(-1))^2)-((n-Y)/(1-p)^2)
199
200 FISMAT=matrix(c(-I11,-I12,-I12,-I22),byrow=T,nrow=2)
201 A=solve(FISMAT)
202 A
203 SE_pie=sqrt(0.00171)
204 SE_lambda=sqrt(0.059)
205 SE_pie
206 SE_lambda
207
208 #-----
209 #-----
210
211         #Computation of SE of MME using Delta Method
212
213 x=0:20
214 f=c(34,5,8,15,16,15,8,6,7,0,3,1,0,0,0,1,0,0,0,0,1)
215
216 n=sum(f)
217 u1=sum(x*f)/sum(f);u1
218 u2=sum((x^2)*f)/sum(f);u2
219 u3=sum((x^3)*f)/sum(f);u3
220 u4=sum((x^4)*f)/sum(f);u4
221
222 var_t1=(u2-u1^2)/n
223 var_t2=(u4-u2^2)/n
224 cov_t1t2=(u3-u1*u2)/n
225
226 d1.dt1=-(u2/u1^2)
227 d1.dt2=1/u1
228
229 dp.dt1=(u1^2-2*u1*u2)/(u2-u1)^2
230 dp.dt2=(u1/(u2-u1))^2
231
232 var_pie.hat=var_t1*(dp.dt1)^2+
233     var_t2*(dp.dt2)^2+
234     2*cov_t1t2*(dp.dt1)*(dp.dt2);var_pie.hat
235
236 SE_piehat=sqrt(var_pie.hat)
237 SE_piehat
238

```

```
239
240 var_lambda.hat=var_t1*(dl.dt1)^2+
241     var_t2*(dl.dt2)^2+
242     2*cov_t1t2*(dl.dt1)*(dl.dt2);var_lambda.hat
243
244 SE_lambdahat=sqrt(var_lambda.hat)
245 SE_lambdahat
246
247 #-----
248 #-----
```