The Pharma Innovation Journal 2021; 10(5): 1471-1473

www.ThePharmaJournal.com

The Pharma Innovation

ISSN (E): 2277- 7695
ISSN (P): 2349-8242
NAAS Rating: 5.23
TPI 2021; 10(5): 1471-1473

www.thepharmajournal.com
Received: 04-03-2021
Accepted: 06-04-2021

TA Raja
Division Agricultural Statistics and Economics, FoA, Sher-e-Kashmir University of Agricultural Science and Technology-Kashmir, Wadura, G.P.O Srinagar, India

S Maqbool
Division Agricultural Statistics and Economics, FoA, Sher-e-Kashmir University of Agricultural Science and Technology-Kashmir, Wadura, G.P.O Srinagar, India

FJ wani
Division Agricultural Statistics and Economics, FoA, Sher-e-Kashmir University of Agricultural Science and Technology-Kashmir, Wadura, G.P.O Srinagar, India

Corresponding Author:
TA Raja
Division Agricultural Statistics and Economics, FoA, Sher-e-Kashmir University of Agricultural Science and Technology-Kashmir, Wadura, G.P.O Srinagar, India

Generalized Poisson distribution and zero inflated Poisson distribution with reference to Poisson distribution

TA Raja, S Maqbool and FJ wani

Abstract
In this paper we study Zero Inflated Poisson Distribution (ZIPD),Generalized Poisson distribution with Poisson Distribution (PD). Goodness of fit for two data sets were carried out by the method of the proportion of the zeroth cell. In first data set ZIPD (p=0.854) with parameters ($w^{\wedge}=0.56$) and ($\lambda^{\wedge}=1.32$) and in second data set ZIPD (p=0.520) with parameters ($w^{\wedge}=0.430$) and ($\lambda^{\wedge}=1.770$). It is very logical and encouraging that Zero Inflated Poisson Distribution (ZIPD) provided better fit as compared to Poisson distribution and Generalized Poisson in both the data sets because of more number of zeros.

Keywords: Poisson distribution, generalized Poisson distribution, zero inflated Poisson distribution and applications

Introduction
Poisson distribution is often used as a standard probability model for data having counts. However there are data sets in which are not well fitted by a Poisson model, because there are more zero counts than are compatible with the Poisson model. Zero inflated models are used to model count data that have many zeros, so inflated Poisson model may be used to model count data for which proportion of zero counts is greater than expected. Zero inflated Poisson model is generally proposed in such situations. Ghosh *et al.* (2006) [3] has rightly found out that when some construction processes are in a straight perfect state, zero or minimum defects will occur with a high probability. However, random changes in the developed situation can lead the process to an imperfect State, producing items with defects. The production process can move randomly. For this type of production process many items will be produced with zero defects and this excess might be better attributed by a zero inflated Poisson model than a Poisson model. Johnson *et al.* (1992) [5] gave a articulate way to model zero inflation. Various authors have considered the zero Inflated Poisson (ZIP) as a possible model for biological count data. An entomological example Desouhant *et al*. (1998) [1] has found that the distribution gave a good fit to 25 out of 31 data sets involving the chestnut weevil. Raja, T A (2012) [8] also worked on zero inflated on Poisson distribution and Raja *et al* (2013) [9] also stated inference of some Poisson type distributions. Researchers assume that the imperfect state is Poisson distributed, which ignores some other useful distributions, such as Negative Binomial Distribution (NBD).

Statement
Assume that the perfect state X_0 is 0 with probability 1 and the imperfect state X_1 is a random variable taking non-negative integers with the following probability density function

$$P\ (X= k) = g\ (k, \lambda) \qquad \ldots\ldots (1.1)$$

For k = 0, 1..., Where $\lambda = (\lambda_1,\lambda_2,\lambda_3,\ldots\ldots\lambda_g)'$ is an unknown parameter vector in an open subset D of S-dimensional Euclidean space R^S. X_i 's distribution is called the distribution of the imperfect state.

Consider the mixture of X_0 *and* X_1 with the distribution Bernoulli (ω) where $0 < \omega \leq 1$. Assume that the probability mass function of the mixture is;

$f\left[k, \binom{w}{\lambda}\right]$

Note that we exclude the case that ω = 0, since that is a trivial case and usually is not interesting. However, we would consider to contain the case that ω = 1, in which only the imperfect state exists. For example, we might be interested in testing that ω = 1, which tests whether the data are from a zero-inflated model or from the distribution of the imperfect state.

Define $\theta = \binom{w}{\lambda}$

and © = (0,1)XD ....(1.2)

we write $f\left(k, \binom{w}{\lambda}\right)$ as f(k, θ) where θ ϵ ©

proposed general zero-inflated models are:-

$f(k,\theta) = (1-w) + wg(0,\lambda)\ for k = 0$ …………(1.3)

and $f(k,\theta) = wg(k,\lambda)\ for k = 1,2,3.$ ………... (1.4)

The equations are known as zero-inflated models with X˜ZIM (θ, g) and X˜ZIM(θ).

Method of proportion of the zero^th cell: Here we compare the observed proportion of zero^th cell and observed mean to their corresponding theoretical values in order to obtain the estimates of the parameters which are given below:

$P_0 = \frac{n0}{N} = 1 - +wg(0, \lambda)$ ……………………………...……….(2.1)

$\mu = wg(y_i, \lambda)$ …………………………………………… ……….(2.2)

where n_0 is the number of observation of zero^th cell and N is the total number of observations, ω and λ are the parameters to be estimated.

Poisson Model

$P(x) = \frac{\lambda^x}{x!} e^{-\lambda}$ ……………………………………….(3.1)

The Poisson model depends upon a single parameter λ, which is mean as well as the variance of the distribution. The equality of mean and variance symbolizes the Poisson distribution among all Power Series Distributions (PSD), Patil(1963).

The estimate of the parameter of PD is as

$\lambda = \bar{x}$ ……………………..……………………..(3.2)

where $\bar{x}$ is the sample mean.

Generalized Poisson Distribution

Consul and Jain (1973) presented a Generalized Poisson distribution (GPD) with two parameters λ_1 and λ_2 and it was obtained by them as a limiting form of another models. Let x be a random variable (r v) defined over a non-negative integral and let $p(\lambda_1, \lambda_2)$ denote the probability that (r v) x takes the non-negative integral values of x. The generalized Poisson distribution (GPD) is defined mathematically by the formula

$$Px(\lambda_1, \lambda_2) = \left\{\frac{\lambda_1(\lambda_1 + x\lambda_2)^{x-1} e^{-\lambda_1 - x\lambda_2}}{x!}\right\}, x = 0,1,2,\ldots \qquad (4.1)$$

0 for x>m, when $\lambda_1 < 0$

and zero other wise where $\lambda_1 > 0$, max $(-1, -\lambda_1/m) \leq \lambda_2 \leq 1$ and $m(\geq 4)$ is the largest positive integer for $\lambda_1 + m\lambda_2 > 0$ when λ_2 is negative.

The parameter λ_2 is independent of λ_1 and the lower limit is imposed to ensure that there are at least five classes with nonzero probability when λ_2 is negative. The symbol λ_1 and λ_2 are called the first and the second parameter of the GPD model. The mean and variance of the GPD is defined as

$$Mean = \frac{\lambda_1}{1 - \lambda_2} \qquad (4.2)$$

$$\text{var}\,iance = \frac{\lambda_1}{(1 - \lambda_2)^3} \qquad (4.3)$$

The variance of this GPD model is greater than, equal to or less than the mean according to whether the second parameter λ_2 is positive, zero or negative and both mean and variance tend to increase or decrease in values, as λ_1 increase or decrease. When λ_1 is positive the mean and variance both increase in values as λ_1 increases but the variance increase faster than the mean. It is obvious that the classical Poission distribution with parameter λ_1 is a special case of our GPD with parameters λ_1 and λ_2 and is obtained when $\lambda_2 = 0$. GPD model is capable of greater changes in variance than mean. GPD has been found useful in various fields like biology, ecology, epidemiology and genetics. Probably due to these reasons it attracted many researchers and so in a very short span of time a number of research papers appeared in the statistical literature. A good account of these can be obtained in Consul (1989). Here it has been found to be a member of Consul and Shenton's (1972) family of Lagrangian distribution and also of the Gupta's (1974) Modified power series distribution (MPSD).

Estimation of Parameters of GPD Model

Consul and Jain (1973) gave the moment estimator's for the parameters of GPD as

$$\lambda_1 = \sqrt{\frac{m_1^{\ 3}}{m_2}} = \sqrt{\frac{x^3}{S^2}} \qquad (4.4)$$

$$\lambda_2 = 1 - \sqrt{\frac{m_1}{m_2}} = 1 - \sqrt{\frac{\bar{x}}{S^2}} \qquad (4.5)$$

where $\bar{x}$ or m_1 and S^2 or m_2 are sample mean and sample variance respectively.

Zero Inflated Poisson Model

$P_0 = 1 - w + we - \lambda$ ………….(5.1)

$Pk = \frac{we-\lambda.}{k1.} \lambda^k$ …………..(5.2)

Method of Proportion of zeroth cell

$x^{-} = w\lambda$ ………………(5.3)

Table 1: Distribution of of observed and expected number of leaves according to the number of insects

No. of Insects (x)	Observed Frequency (f)	PD	GPD	ZIPD
0	33	26	28	33
1	12	20	18	11
2	6	8	7	7
3	3	2	2	4
4	1	0	1	1
5	1	0	0	0
Total	56	56.0	56.0	56.0
Estimation of parameters	W^ λ^ Λ_1 Λ_2	0.75	0.474 0.785	0.56 1.32
p-value	-	0.452	0.736	0.854

In this distribution the values of w and λ were found to be 0.56, and 1.32, for zero inflated Poisson distribution(ZIPD) as per method proportion of the zeroth cell and value of λ for Poisson distribution is 0.75, value of $\lambda 1$ and $\lambda 2$ 0.474 and 0.785 Further p-value reveals that ZIPD provides a good fitting 0.854 against 0.452 and 0.736.

Table 2: Distribution of observed and expected number of redmites on leaves.

No. of mites on leaves (x)	Observed Frequency (f)	PD	GPD	ZIPD
0	37	28	32	40
1	12	23	19	9
2	8	10	9	7
3	5	3	4	4
4	2	1	1	3
5	1	0	0	2
	65	65.0	65.0	65.0
Estimation of parameters	W^ λ^ Λ_1 Λ_2	0.86	0.562 0.766	0.43 1.77
p-value	-	0.18	0.47	0.52

In this distribution the values of w *and* λ are found to be 0.43 and 1.77 and λ for zero inflated Poisson distribution (ZIPD) as per method proportion of the zeroth cell and λ for Poisson distribution is 0.86 **,** $\lambda 1$ and $\lambda 2$ *0.562* and *0.766***.** Further p-value reveals that ZIPD provides a good fitting 0.52 against 0.18 and 0.47.

Result & Discussion

It is encouraging to observe that a close fit was observed in case of Zero Inflated Poisson Distribution in comparison to Poisson distribution and Generalized Poisson distribution in both the applications. In table1 it is obvious that 59% observations were with no insects and p-value observed via ZIPD was quite satisfactory 0.854. In table 2, 57% observations were with no mite and p-value via ZIPD was 0.520. The degree of insignificance was higher in zero Inflated Poisson Distribution than Poisson and Generalized Poisson distribution.

References

1. Desonhant E, Debouzie D, Menu F. Oviposition pattern of Phytophagous insects: On the importance of host population heterogeneity. Oecologia 1998;114:382-388.
2. Dianewith Lmbert. Zero Inflated Poisson regression, with an application to defects in manufacturing Technometrics 1992;34(1):45-55.
3. Gosh JK, Delam Pady M, Samanta T. An introduction to Bayesian analysis theory and methods. Springer texts in statistics, Springer New York 2006.
4. Heilbron DC. Zero-altered and other regressions models for counts data with added zeros. Biometric. J. 1994;36:531-547.
5. Johnson N, Kotz S, Kemp AW. Univariate Discrete Distributions, second edition, Wiley New York 1992.
6. Lambert D. Zero-inflated Poisson regression, with an application to defects in manufacturing. Technometrics, 1992;34:1-14.
7. Patil GP. Certain properties of the generalized power series distribution. Ann. instt. Math 1963;14:179-182.
8. Raja TA. On Zero Inflated Poisson Distribution with Reference to Poisson Distribution. Int. J. Modern Math. Sci 2012;3(2):85-90
9. Raja TA, Agnihotri RK. On Inference of Some Poisson Type Distributions With Applications. Int. J. Agricult. Stat. Sci. supplement 2013;9(1):113-118.