On Applications of Some Probability Distributions

T. A. Raja and A. H. Mir

Division Agricultural Statistics, Sher-e- Kashmir University of Agricultural Science and Technology-Kashmir, Shalimar-191121, G.P.O Srinagar-190001

ABSTRACT

The paper pertains to the inferences of Poisson, Poisson-Lindly, Generalized Poisson, Negative Binomial and Generalized Negative Binomial distributions. Parameter estimation is performed and goodness of fit for two data sets are obtained and some new applications are tried. The first set is ecological data, where distribution of 102 spiders under 240 boards (Cole, 1946) and the second set is biological data pertaining to distribution of infections of 31 infants with *Plasmodium falciparum* (malaria, the killer parasite) (Molineaux and Gramiccia, 1980) is considered. Generalized Poisson distribution and Generalized Negative Binomial distribution result in better fitting and infer their preeminence.

Key words: Probability distribution, parameter estimation, applications

Poisson Distribution

French mathematician Simon Denis Poisson (1781-1840) derived the distribution. Let X be a discrete random variable (r, v) defined over non-negative integral values and let Px(λ) denote the probability that the random variable x takes the non-negative integral values X. The Poisson distribution (PD) is defined mathematically by the formula.

$$p(x \setminus \lambda) = \begin{cases} \dfrac{\lambda^{x} e^{-\lambda}}{x!}, x = 0,1,2 \ldots\ldots \end{cases}$$

0, otherwise

Where (λ >0) and is known as parameter of the distribution. It is generated by processes in which a large number of cells, squares, leaves, petals or intervals of time (e.g. seconds, minutes, hours, and days) are hit by a relatively small number of events (births, deaths, blood cells, nuclear decay particles, ball spot, etc), Such that the occurrence or non occurrence of events in an interval and the probability of two or more occurrences of events in an interval of time is almost zero i.e. a cell (or interval) with certain counts is likely to get another count as a cell with fewer counts or with no counts at all. This principle of randomness implies that the individual organisms or events are scattered by chance alone.

The Poisson model depends upon a single parameter λ, which is mean as well as the variance of the distribution. The principle of complete randomness in Poisson distribution is excellent but it is not practical for all situations. (Consul, 1989).

Estimation of parameters of Poisson Distribution

Simon Denis Poisson (1781-1840) gave the moment estimator for the parameter of PD as:

$$\lambda = \bar{x} = m1$$

Where λ or m1, is the sample mean.

Poisson-Lindley Distribution

Lindley (1958) derived a distribution known as Lindley distribution based on Baye's Theorem. Sankaran (1970) generalized Lindley distribution by mixing with Poisson distribution which is known as Poisson- Lindley distribution. The probability generating function of Poisson- Lindley distribution obtained by compounding the Poisson distribution with one due to Lindley may be defined as

$$g(t) = \frac{(\theta + 2 - z)\theta^2}{(\theta + 1)(\theta + 1 - z)^2}, \qquad \theta > 0.$$

The probability recurrence relation may be expressed as

$$P_{r+1} = \frac{P_r(\theta + 3 + r)}{(\theta + 1)(\theta + 2 + r)}, \qquad \text{for } r \geq 0,$$

where $P_0 = \frac{\theta^2(\theta+2)}{(\theta+1)^3}$. Here $\mu = \frac{\theta+2}{\theta(\theta+1)}$, and

$$\sigma^2 = \frac{\theta^3 + 4\theta^2 + 6\theta + 2}{\theta^2(\theta+1)^2},$$

be respectively the mean and variance of the distribution.

Estimation of parameters of Poisson-Lindley Distribution

The estimation of the parameter θ of the distribution is related to μ by

$$\theta = \frac{-(\mu-1)+\sqrt{(\mu-1)^2+8\mu}}{2\mu},$$

Generalized Poisson Distribution

It was derived by Consul and Jain in (1973).They defined a generalized Poisson distribution (GPD) with two parameters λ_1 and λ_2 as

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

0 for x>m, when λ_1 <0

and zero otherwise where $\lambda_1 > 0$ max (-1, λ_1/m)$\leq \lambda_2 \leq 1$ and m $\leq$ 4 is the largest positive integer for $\lambda_1 + m\,\lambda_2 > 0$ where λ_2 is negative. The parameter λ_2 is independent of $\lambda 1$ and the lower limit is imposed to ensure that there are at least five classes with non-zero 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 author's defined mean and variance of the GPD as

Mean = $\frac{\lambda_1}{1-\lambda_2}$

and variance = $\frac{\lambda_1}{(1-\lambda_2)^3}$

The variance of this GPD model is greater than, equal 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.

Although the GPD model has only two parameters λ_1 and λ_2 it provides excellent fit to various types of observed data. *Janardan* and *Schacffer* (1977) have applied the GPD models to about 100 different sets of biological mechanisms.

GPD has been found useful in various fields like biology, queuing theory, 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 this can be obtained in Consul (1989). Here it has been found to be a member of *Consul* and *Shenton's* (1972) family of Lagrangian distributions and also of the Gupta (1974) modified power series distributions (MPSD).

Estimation of parameters of Generalized Poisson Distribution.

Consul and Jain (1973) obtained by moment method estimators for the parameter of GPD in the form as:-

$$\lambda_1 = \sqrt{\frac{m_1^3}{m_2}} = \sqrt{\frac{\bar{x}^3}{s^2}}$$

$$\lambda_2 = 1 - \sqrt{\frac{m_1}{m_2}} = 1 - \sqrt{\frac{\bar{x}}{s^2}}$$

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

Negative Binomial Distribution (NBD)

Negative Binomial distribution (NBD) was discussed by Pascal first in 1914 for getting r success and x failures in exactly (r + x) independent trails, where X is a random variable. The probability mass function of the Negative Binomial distribution (NBD) is given by

$$P(X = x) = \binom{x + r - 1}{r - 1} p^r q^x \quad , \quad x = 0,1,2,\ldots \; 0 < p < 1$$

The mean and variance of NBD are rq/p and rq/p^2. The NBD provides a good fit to the situation where the mean is always less than the variance. It is the most widely used distribution among the two parameter family of distributions and also found its application in the biological cases. (Bliss and Fisher, 1953) was the first to NBD to biological data.

Estimation of Parameters of Negative Binomial Distribution

The estimation of the parameters of NBD by the method of moments given as

$$\hat{p} = \frac{\mu_1'}{\mu_2}\ , \quad \text{where}\ \mu_1' = \frac{\Sigma fx}{\Sigma f}\ ,\ \mu_2 = \frac{\Sigma fx^2}{\Sigma f} - \left(\frac{\Sigma fx}{\Sigma f}\right)^2$$

$r = \dfrac{p\mu_1'}{q}$, where q = 1 – p

Generalized Negative Binomial Distribution

Jain and *Consul* (1971) defined the Generalized Negative Binomial distribution (GNBD) given by its probability function as:

$$P(x = x) = \frac{n}{n+\beta x}\binom{n+\beta x}{x}\alpha^{x}(1-\alpha)^{n+\beta x - x}$$

= 0 for x ≥ m such that $n + \beta x < 0$

= 0 < α< | ; n> 0 and | αβ | < |

It is also known as Lagrangian Negative Binomial distribution. The probability model reduces to the binomial distribution when β=0 and m is an integer, and to the negative binomial distribution when β = 1. it also resembles to the Poisson distribution at β = ½ as for this value of β the mean and the variance are approximately equal.

The GNBD is also a member of the Gupta (1974) modified power series distribution (MPSD). It is a member of *Consul* and *Shenton* (1972) Lagrangian probability distribution. *Jain* and *Consul* (1971) obtained the first four moments of GNBD and discussed its various properties

The GNBD model has many important applications in various field of study and is useful in queuing theory and branching process. It has an important use in chemistry in the reaction called polymerization. *Famoye* and *Consul* (1989) considered a stochastic urn model for the GNBD and gave some other interesting application of this model.

Gupta (1975) discussed the ML estimation of the GNBD as a particular case of the MPSD. The estimation of parameters by maximum likelihood method becomes tedious as the three likelihood equation come in such a form that don't seem to be easily solvable. However they can be solved using some interaction techniques. But as the moments of GNBD come in the forms, the method of moments of estimation can be used easily.

Estimation of Parameter of GNBD Model

Jain and *Consul* (1971) used the method of moments to estimate the GNBD parameters. The first three moments of the distribution namely the mean (μ_1), the variance (μ_2) and the third central moment (μ_3) are respectively

$$\mu_1 = \frac{n\alpha}{(1-\alpha\beta)}$$

$$\mu_2 = \frac{n\alpha(1-\alpha)}{(1-\alpha\beta)^3}$$

and $$\mu_3 = \frac{n\alpha(1-\alpha)}{(1-\alpha\beta)^5}\left[1-2\alpha+\alpha\beta(2-\alpha)\right]$$

From these equations after substitution, we get

$$n = \mu_1' \left(1-\alpha\beta\right)/\alpha$$

$$\beta = \frac{1}{\alpha}\left[1-\left(\frac{m_1(1-\alpha)^{1/2}}{m_2}\right)\right]$$

the roots will be real if K ≥ 4. the value of root lying between 0 and 1 is, therefore given by

$$\alpha = 2-\frac{k}{2}+\sqrt{\left(\frac{K^2}{4}-K\right)}$$

the three moments μ_1 μ_2 and μ_3 are replaced by their estimate from the frequency distribution of a sample as

$$\mu_1 = \frac{\sum fx}{N}, N = \sum f$$

$$\mu_2 = \frac{N(\sum f^2 x) - (\sum fx)^2}{N(N-1)}, N = \sum f$$

$$\mu_3 = \frac{\sum f^3 x - 3\mu_1\left(\sum fx^2\right) + 2N\mu_1}{N}, N = \sum f$$

From the above equation, the estimates of $\alpha_, \beta$ and n can be obtained

Goodness of Fit

The Poisson distribution is used in biological and ecological problems. The Poisson model has been used in very wide range of situations to describe the behavior of living beings. The principal of complete randomness in Poisson distribution is excellent but is not practical for all situations. Various examples can be provided from biological and ecological system where this principle of complete randomness is not sufficient to describe the behavior of livings beings. The Poisson –Lindley, and Generalized Poisson distribution being the generalization of the Poisson distribution can give a reasonably good fit to the data. Because of the versatility of GNBD and GPD lot of work has been done on it.

Firstly here we have tried distribution of 102 spiders under 240 boards. (data of Cole, 1946) . (Table 1)

Table 1:- Distribution of 102 Spiders under 240 boards.

No. of Spiders Per board	**Observed freq**	**PD**	**PLD**	**GPD**	**NBD**	**GNBD**
0	159	157.2	158.3	159.0	156.4	158.6
1	64	66.5	66.3	63.4	67.4	65.2
2	13	14.2	12.8	14.6	12.2	12.6
3	4	2.0	2.6	3.0	3.0	13.5
Total	240	240	240	240	240	240

Parameters	λ=0.425	θ=1.620	λ_1=.4114	p=0.754	α=0.731
			λ_2 =.0320	r=1.305	β=0.970
					n=0.245
χ^2 =	0.785	0 .384	0.294	0.378	0.186
P-value	0.853	0 .924	0 .961	0.921	0.978

The second data which we tried has been collected by the World Health Organization (WHO) in Garki Nigeria (Molineaux and Gramiccia, 1980) the data of infection with *Plasmodium falciparum* (Malaria the killer parasite).The distributions has been fitted to the data as. (Table. 2)

Table 2:- Distribution of infections of 31 infants with Plasmodium falciparum (malaria, the killer parasite).

No. of Infections	Observed freq	PD	PLD	GPD	NBD	GNBD
0	7	3.10	3.15	5.12	4.29	4.89
1	3	7.14	7.00	6.23	7.38	6.22
2	5	8.24	8.12	7.14	7.26	7.10
3	10	6.33	6.45	9.10	5.37	6.54
4	4	3.65	4.14	7.05	3.31	3.85
5	1	1.70	0.97	0.45	1.80	1.90
6	0	0.60	0.62	0.09	0.89	0.85
7	1	0.24	0.12	0.58	0.51	0.68
Total	**31**	**31**	**31**	**31**	**31**	**31**

Parameters	λ=2.308	θ=1.352	λ_1=.584	p=0.287	α=0.625
			λ_2 =.142	r=8.10	β=0.895
					n=0.278
χ^2 =	14.65	10.55	7.56	9.80	8.24
P-value	0.043	0.095	0.268	0.235	0.187

The above applications infer that Generalized Poisson distribution and Generalized Negative binomial distribution are applicable in a wider range of varieties of observed biological and ecological data. Thus it seems to be very logical that GPD and GNBD give comparatively better fits to the data. In table1 the estimate of the parameters in Poisson(λ=0.425), P Lindly (θ=1.620) , GPD(λ_1=.4114, λ_2 =.0320), NBD(p=0.754, r=1.305) and GNBD (α=0.731, β=0.970, n=0.245) and in table2 the estimate of the parameters in Poisson(λ=2.308), P- Lindly (θ=1.352) and GPD(λ_1=.584, λ_2 =.142) , NBD(p=0.287, r=8.10) and GNBD (α=0.625, β=0.895, n=0. 278) The computed value of Chi-Square (χ^2) reflected by their p-value further show that non-significant difference occur in table 1 throughout but with varying degree of non –significance slightly higher in GNBD, and followed by GPD . In table2 the computed value of Chi-square(χ^2) reflected by their p-value is significant for PD and PLD and non-significant in GNBD, NBD and GPD. The better fitting by GNBD and GPD infer their superiority over other distributions.

REFERENCES

Bliss, C.I and Fisher R.A.1953.: Fitting of the Negative binomial distribution to biological data and note on the efficient fitting of the binomial. *Biometrics*, **9**: 176-200.

Cole, L.C.1946. A theory for analyzing contagiously distributed Populations. *Ecology*, **27**(4),329-249.

Consul, P.C. 1989. *Generalized Poisson distribution. properties and applications*. Marcel Dekker Inc: New York.

Consul, P.C. and Jain, G.C.1973. A generalization of the Poisson distribution, *Technometrics*,**15**(4): 791-799

Consul, P.C. and Shenton, L. R.1972. use of lagrang expansion for generating Generalized Poisson distribution SIAMS*. J. Appl. Math,* **23**(2):239-249.

Famoye, F. and Consul, P C.1989. Confidence. Internal estimation in the class of modified power series distribution. *Statistics*, **20**(1): 141-148.

Gupta, R. C.1974. Maximum likelihood estimation of a MPSD and its applications. *Sankhya, Series* B, **36**(3):288-298.

Janardan, K. G. and Schacffer. D. J. 1977. Models for the analysis of chromosomal aberrations in human leukocytes. *Biometrical J.* **19**(8) : 595- 612 .

Jain, G. C. and Consul. P. C.1971. A generalized negative binomial distribution. *SIAM. J. Appl,* **21**(4): 501-513.

Lindley, D. V.1958. Fiducial distribution and Bayes theorem, *Journal of the Royal Statistical Society*, **20:** 102-107.

Molineaux, L. and Gramiccia, G. 1980.*The Garki Project*, World Health Organization.

Patil, G.P.1963. Certain Properties of the generalized Power Series Distribution, *Ann. Instt. Math,* **14**:179-182.

Sankaran, M.1970.The Discrete Poisson-Lindley distribution. *Biometrics,* **26:** 145-149.