GLOBAL JOURNAL OF PURE AND APPLIED SCIENCES, VOL.7, NO. 4, SEPTEMBER 2001: 749-753
 ISSN 1118-0579

EXPLORING THE POTENTIALS OF THE NEGATIVE BINOMIAL DISTRIBUTION

A. C. EKE, B.O. EKPENYONG, and G. I OGBAN

(Received 24 March 2000 ; Revision accepted 6 July, 2001)

ABSTRACT

More than ever, the applications of the negative binomial distribution have received greater attention especially in agriculture, biology, health and accident statistics, population counts and communication networks. This paper sets out to review the forms and areas of applications of the negative binomial distribution (NBD). An example is presented.

Key Words : Poisson distribution, Negative binomial, Reversed-J, contagious distribution, truncated form.

INTRODUCTION

The negative binomial distribution becomes necessary for use and application when the poisson distribution fails in both its assumptions and applications. It is to be recognised that the poisson distribution describes certain cases in cosmic radiation in physics, incidences of strikes, break downs in telephone equipment and certain other forms of "accidents." However, in many situations the simple conditions for the poisson distributions that successive events occur independently of each at a constant rate do not quite hold. Consequently, irregular events which do not appear to follow simple independent probability processes have to be described by more complex stochastic models. Infact, the simplicity of the poisson distribution with a single parameter makes it rather inflexible. Indeed the main importance of the poisson distribution is a constituent of more complex models.

SIX (6) CONFIGURATIONS OF THE NBD

The NBD is either reverse-J-shaped or if the mean is large, humpbacked with a long positive tail to the right.

THE PASCAL FORM (FELLER, 1957)

This is given by

$$P(i) = \binom{i+r-1}{i} P^r q^i \quad (i = 0, 1, 2 \ldots) \quad (r = 1,2,3) \qquad \ldots\ldots (1)$$

Here p may be estimated if unknown but r is usually known. Haldane (1943) has provided an unbiased estimator of p. Both the approximate and exact expressions for the variance of p have been provided by Haldane (1943) and Finney (1949) respectively.

THE ANSCOMBE FORM (ANSCOMBE, 1950)

Here the NBD depends on two parameters (m,k). With m as the mean and k as the variance , the form is

$$P(i) = \frac{\Gamma(k+1)}{i!\,\Gamma(k)} \left(1+\frac{m}{k}\right)^{-k} \left(\frac{m}{m+k}\right)^{i}, \; i = 0,1,2\ldots\ldots m > o \qquad \ldots\ldots (2)$$

The parameter k is positive but usually non-integral which can be estimated by the maximum like hood method.

A. C. EKE, Died November, 2000. Department of Statistics, The Polytechnic, Calabar
B. O. EKPENYONG, Department of Statistics, The Polytechnic, Calabar
G. I. OGBAN, Department of Mathematics , The Polytechnic, Calabar

THE TRUNCATED FORM

Rider (1955) and Sampford (1955) have observed that in some cases, the value of i = o cannot be observed . If the NBD of equation (1) is divided by 1-p(o), where p(0) = p^r, we have

$$P(i)=\binom{i+r-1}{i}P^r q^i(1-p^r)^{-1},(i=1,2....) \quad (3)$$

Sampford (1955) has an iterative method of obtaining the moment and maximum likelihood estimators of the parameters of the truncated NBD. Brass (1958) has simplified methods for estimating those parameters.

THE OLDHAM FORM (OLDHAM, 1968)

Oldham used

a = k/m, b = k to obtain

$$P(i)=\binom{b+i-1}{i}\left(\frac{a}{a+1}\right)^b\left(\frac{1}{a+i}\right)^i,(i=0,1;2.....) \quad (4)$$

The estimation of the parameters in the NBD Oldham form has been discussed by Fisher (1941) and Bliss and Fisher (1953).

The Fisher Form. Fisher (1941)

Fisher used

$P = \frac{m}{k}$, q = 1 + P, $P_0 = q^{-k}$ to obtain

$$P(i)=p_0\binom{i+k-1}{i}\left(\frac{p}{q}\right)^i,(i=0.1,2.....) \quad (5)$$

In this notation, the probability generating function (pgf) is $(q-pt)^{-k}$, and the mean, variances and corrected third moment are kp, kpq and kpq (q+p) Ross and Preece (1985). These results illustrate the anology with the ordinary binomial distribution, whose probability generating function is $(q+pt)^n$ and whose mean, variance and corrected third moment are np, npq and npq (q-p), with q=1-p.

<u>**The Bliss Form**</u> (Bliss,1953)

Bliss used Fisher's p and q, with R=p/q which gives,

$$P(i)=\binom{i+k-1}{i}(R)^i,(i=0.1,2.....) \quad (6)$$

<u>**The Freeman Form (Freeman,1980)**</u>

Freeman used Fisher's p and also λ = m which gives

$$P(i)=\binom{i+k-1}{i}\left(\frac{\lambda}{k+\lambda}\right)^i,(i=0,1,2....) \quad (7)$$

<u>**USES OF THE NBD**</u>

The NBD may arise as a result of randomly distributed colonies. If colonies or groups of individuals are distributed randomly over an area (or in time) so that the number of colonies observed in samples of fixed area (or duration) has a poisson distribution, we obtain a negative binomial distribution for the total count if the number of individuals in the colonies are distributed independently in a logarithmic series distribution. Queenouille (1949), williamson and Brether Ion (1964), have discussed the model of randomly distributed colonies in the context of industrial purchasing. Neyman (1939), Jones et al (1984), Anscombe (1949), Skellam (1952), and Evans (1953) have applied the model in entomology and bacterilogy.

McKendrick (1914), Yule (1924), Furry (1939) and Kendall (1949) have discussed a model of population growth, in which there are constant rates of birth and death per individual and a constant rate of immigration which is NBD. This model has been applied to the growth of some living populations e.g population of bacteria and to the spread of an infectious disease in a community (Irwin, 1954). Other situations where NBD has found a reasonable fit can be found as discussed by wise (1984), Oakland (1980), Magee (1953), Brass (1958) and Kemsley (1965).

Yule (1910) and Haldane (1945) have shown that if a proportion p of an individual in a population possess a certain characteristic, the number of observations in excess of r which must be taken to obtain exactly individuals with the characteristic has the Paschal form of the NBD. The Anscombe form given above is a model which is widely applied to accident statistics (Green wood and Yule, 1920; Arbous and Kerich, 1951; Adelstein, 1952; Grosswell and Froggart 1960; Irwin, 1964). The compound poisson model of the NBD has been applied to the study of consumer purchasing behaviour (Ehrenberg, 1959; Chatfield and associates, 1966; Chaffield and Goodhardt, 1970; Charlton and associate, 1972).

APPLICATION OF THE NBD

We shall now consider the application of the NBD.Consider the following (hypothetical) problem

Twenty-five leaves were selected at random from each of six similar apple trees in an orchard, and the adult female European red mites on each leaf counted:

No. of mites per leaf	0	1	2	3	4	5	6	7	8
No. of leaves (frequency)	70	38	17	10	9	3	2	1	0

source : Annual Abstract of statistics 1972.

We desire to fit a negative binomial distribution to the data.

Result

Chi-squared = 4.22 on 6 d . f

Parameter

1	Mean	1.1669	0.12729
2.	Variance	2.43008	0.53786
3.	K	1.02456	0.27584
4.	1/K	19.97603	0.26278

No of mites:	0	1	2	3	4	5	6	7	8
fo:	70	38	17	10	9	3	2	1	0
fe:	69.5	37.6	20.1	10.7	5.7	3.0	1.6	0.8	1.0

COMMENTS

Obviously the poisson distribution would not have fitted the above data properly since the mean is not equal to the variance, the negative binomial distribution of the Anscombe form is appropriate here.

CONCLUSION

The various mathematical models of the NBD have been highlighted. The six model forms discussed he.e can be adapted to different situations and areas. The NBD has the potentials of overcoming the simplistic nature of the poisson distribution. An additional parameter for the NBD gives it an added advantage over the poisson.

REFERENCES

Adelstein, A. M 1952. Accident proneness a criticism of the concept based upon an analysis of shunters' accidents J. R. Statist. Soc., A 115: 354-410

Anscombe, F. J.1949. The Statistical analysis of insect counts based on the negative binomial distribution. Biometrics, 5:165-175

Anscombe, F. J. 1950. Sampling theory of the negative binomial and logarithmic series distributions. Biometrica, 37,358-382.

Arbous, A. G. & Kerrich, J. E. 195.1 Accident statistics and the concept of accident proneness Biometrics, 7: 340-342

Bliss, C. I. 1953 Fitting the negative binomial distribution to biological data (with Note on the efficient fitting of the negative binomial by R. A. Fisher), Biometrics,9: 17-200.

Brass, N. 1958 Simplified methods of fitting the truncated negative binomial distribution. Biometrica, 45. 59-68.

Charlton, p ., Ehrenberg, A. S. C. & Pymont, B. 1972 Bayer behaviour under mini-test conditions . J. Market Res. Soc.14:171-183.

Chatfield, C. 1969 on estimating the parameter of the logarithmic series and negative binomial distributions.Biometrica, 56: 411-414

Chartfield, C. & Goodhardt, G. J .1970. The beta-binomial model for consumer behaviour. Appl. Statist., 19: 240-250

Chatfield, c., Ehrenberg, A. S. C. & Goodhardt, G. J. 1966 Progress on a simplified model on stationary purchasing behaviour . J. R. Statist. Soc. A, 129, 317-367.

Crosswell, W. L. & Froggoart, P. 1963 Causation of bus driver accidents-an epidemiologica study. London: Oxford University press.

Ehrenberg, A. S. C. 1959 The pattern of consumer purchases Appl. Static., 8: 26-41

Evans, D. A. 1953. Experimental evidence concerning contagious distribution in ecology. Biometrika, 40: 186-211.

Feller, W . 1943. O n a general class of contagious distributions. Ann. Maths. Statist., 14: 389-400.

Feller W. 1957. An introduction to probability theory and its applicaions Wiley, New York.

Finney, D. J. 1949. On a method of estimating frequencies Biometrika, 36: 233 - 235.

Fisher, R. A. 1941. The negative binomial distribution Annals of Eugenics, 11: 182-1987. (Reprinted as paper 38 in: Fisher, R. A. 1950. Contributions to Mathematical Statistics (New York, Wiley).

Fisher, R. A. Corbett, R. S. & Williams, C. B. 1943. The relation between the number of species and the number of individuals in a random sample of an animal population. J. Animal Ecology 12: 42-58.

Freman, G. H. 1980. Fitting two-parameter discrete distributions to many data sets with the common parameter, Applied Statistics, 29: 259-267.

Furry, W. H. 1939. On fluctuation phenomena in the passage of high energy electrons through lead. Phys. Rev., 52: 569 - 581.

Grainger, R. M. & Reid, D. B. W. 1954. Distribution of dental cares in children. J. Dent. Res. 33: 613 -623.

Geig-Smith, P. 1952. Quantitative plant ecology. London, Butterworth.

Geig-Smith P. 1983. Quantitative plant ecology, 3rd ed. Studies in Ecology, 9 (Oxford, Blackwell Scientific).

Greenwood, M & Yuke. G. U. 1920. An inquiry into the nature of frequency distributions of multiple happenings. J. R Statist. Soc. 8: 255 -279.

Haldane, J. B. S. 1941. The fitting of binomial distributions. Annals Eugenices, 11: 179-181.

Haldane. J. B. S. 1943) On a method of estimating frequencies. Biometrika, 3: 222- 225.

Haldane, J. D. S. 1945. A labour saving method of sampling. Nature, 155: 49-50.

Irwin, J. D. 1964. The personal factor in accidents a review article. J. R. Statist. Soc. A. 127: 438-451.

Jones, P.C.T., Mollison, J. E., & Ouenoullie, M. H. 1948. A technique for the quantitative estimation of soil micro-organisms. J. Gênetal Microbiology, 2: 54-69,

Kensley, W. F.F. 1965. Interview variablity in expenditure surveys. J.R Statist. Soc. A. 12: 438-451

Knedall, D.G. 1949. Stochastic processes and population growth. J. R. Statist Soc. B,11: 230-264.

Macgee J. F. 1953. The effect of promotional effort on sales. Opens Res.; 1: 64-74.

Mckundrick A. G. 1914. Studies on the theory of continuous probabilities with special reference to its bearing on natural phenonmona of a progressive nature Pore London Math: Soc. 2(13): 401-416.

Neyman, J. 1939. On a new class of contagious distributions applicable in entomology and bacteriology Ann. Math. Statist., 10: 35-37.

Ofosu, J. B. 1986. The negative binomial distribution: A review of properties and applications. The Nigeria State. Asso. 3 1,8-16 oakland G. B. 1950. An application of sequential analysis to whiltefish sampling. Biometrics, 6: 59-67.

Oldham, P. D. 1968. Measurement in Medicine: The interpretation of numerical data. London English Universities press.

Paerson, K. & Fieller, E. C. 1933. On the applications of the Double bessel function to statistical problems. Biometrika. 25: 158-178.

Patil, G P. 1960. On evaluation of negative binomial distribution function. Ann Math. Statist; 31: 527.

Patil, G. P. 1962. Some methods of estimation for the logarithmic series distribution. Biometrics, 18: 68-75.

Quenoville, M. H. 1949. A relation between the logarithmic Poisson and negative binomial series. Biometrics, 5: 165-164.

Rider, P. R. 1955. Truncated binomial and negative binomial distributions. J. Amer. Assoc., 50: 877-883.

Sampford, M. R. 1953. The truncated negative binomial distribution, Biometrika; 42: 58-69.

Skellam, J. G. 1952. Studies in Statistical Ecology 1, Special Pattern Biometrika, 39: 346-362.

Waters, W. E. 1955. Sequential analysis of forest insect surveys. Forest Science, 1: 68-79.

Williamson. E. & Brethreton, M. H. 1964. Table of the logarithmic series distribution. Ann. Math. Statist, 35: 284-97.

Wise, M. E. 1948. The use of the negative binomial distribution in an industrial sampling problem, J. R. Statist. Soc. B., 8: 202-221.

Yule, G. U. 1910. On the distribution of deaths with age when the causes of death act cumulatively and similar frequency distributions. J. R. Statist. Soc., 73, 26-38.

Yule, G. U. 1924. A mathematical theory of evolutions, based on the conclusions of Dr. T. C. Wills. Hill Trans. Roy. Soc. London, B. 213: 21-87.