A New Generalization of the Negative Binomial Distribution

Ramesh C. Gupta
Department of Mathematics and Statistics
University of Maine
Orono, ME 04469-5752
U.S.A.
rcgupta@maine.maine.edu

S.H. Ong
Institute of Mathematical Sciences
University of Malaya
50603 Kuala Lumpur
Malaysia
ong@omega.math.um.edu.my

1. Introduction

It is well known that the negative binomial distribution (NBD) has become increasingly popular as a more flexible alternative to the Poisson distribution especially when it is doubtful whether the strict requirements particularly independence for a Poisson distribution will be satisfied. For various applications of NBD, see Johnson et. al. (1992).

In this paper we propose a new generalization of the negative binomial distribution which which may be formulated as a mixed Poisson distribution with a generalized gamma (Armero and Bayarri, 1993); Agarwal and Kalla, 1996) as mixing distribution. The resulting distribution has been fitted to various data sets and exhibits a better alternative to the NBD and some of its extensions. Rao's score test is developed for testing the NBD versus the proposed model and is applied to two data sets from the literature. Finally, simulation studies are carried out to investigate the relative error incurred if NBD is used instead of the proposed model.

2. The Proposed Model

The new generalization of the negative binomial distribution has probability mass function (pmf)

$$P(k) = \frac{(m)_k}{k\,!}\, n^k \,\frac{\varphi\left(m+k,\, m+k-\lambda+1;\ (\alpha+1)n\right)}{\varphi\left(m,\, m-\lambda+1;\ \alpha n\right)}$$

where

$$\varphi\ (a, c; x) = \frac{1}{\Gamma(a)} \int_0^\infty \frac{e^{-xt}\ t^{a-1}}{(1+t)^{a-c+1}}\ dt\,, \qquad c > a > 0$$

is the confluent hypergeometric function of the second kind (Erdélyi, 1953, vol. I, page 255).
Note that λ =0 gives the NBD while $\alpha \to 0$ with $q = \lambda - m$ $(\lambda > m)$, $p = m$ and $b = n^{-1}$ gives a generalization of the NBD of Ong (1995). A recurrence relation between the probabilities is given by

$$\begin{aligned}(\alpha+1)(k+1)\ k\,P(k+1) &= k\left(k-\lambda+m-(\alpha+1)\,n\right)P(k)\\ &\quad + (m+k-1)\,n\,P(k-1)\,,\ \ k \geq 1\end{aligned}$$

The probability generating function of the new generalization of the NB is

$$G(z) = \left(1+\alpha^{-1}-z/\alpha\right)^{l-m}\ \frac{\Gamma_l\left(m, \alpha\left(1+\alpha^{-1}-z/\alpha\right)\right)}{\Gamma_l(m, \alpha\, n)}$$

where $z < \alpha + 1$ and the factorial moments are $m_{(r)} = \alpha^{-r}\ \Gamma_l\left(m + r,\ n\right) / \ \Gamma_l\left(m, \alpha\right)$

$$\Gamma_l\ (m,\ \alpha\ n) = \ \Gamma(m)\,/(\alpha\, n)^{l-m}\ \varphi\ (m,\ m - l\ + 1\ ; \alpha\ n)$$

3. Parameter Estimation, Rao' s Score Test and Error Study

Maximum likelihood (ML) estimation is considered for the parameter estimates of the proposed generalization of the NB distribution. Due to the complicated likelihood function, the ML estimates are determined by a direct numerical search of the log-likelihood surface.

Two data sets are examined: (a) quadrat counts of Liatris aspera in an old field association (Table 5, Bliss, 1953) and (b) the number of European red mites on apple leaves (Table1, Bliss, 1952). The fitting of the proposed model is compared with the NBD and a generalized NBD (GNBD) of Jain and Consul (1971). In both the data sets, the proposed model fits well.

To test the hypothesis

$H_0\ : \lambda\ = 0,\ m, n, \alpha$ *unspecified* versus $H_A\ : \lambda\ \neq 0, m, n, \alpha$ *unspecified*

Rao's score test is considered. Partial derivatives of the probabilities have been derived in order to implement the score test. For the two data sets considered, the score test rejects the null hypothesis at the 5 % level.

Simulation studies have been carried out to investigate the relative error incurred if NBD is used instead of the proposed model for various values of λ keeping the other parameters fixed. It is observed that the relative errors increase with increasing λ.

REFERENCE

Agarwall, S.K. and Kalla, S.L. (1996). A generalized gamma distribution and its application in reliability. Commun. Statist. Theor. Meth., 25 (1), 201 – 210.

Armero, C. and Bayarri, M.J. (1993). A Bayesian analysis of a queueing system with unlimited service, Technical Report No. 93-50, Department of Statistics, Purdue University, West Layette, IN

Bliss, C.I. (1953), Fitting the negative binomial distribution to biological data. Biometrics, 9, 177 – 200.

Erdélyi, A. et al (1953), Higher Transcendental Functions. McGraw-Hill, New York

Jain, G.C and Consul, P.C. (1971). A generalized negative binomial distribution. SIAM Journal of Applied Mathematics, 21, 501 – 513.

Johnson, N.L., Kotz, S. and Kemp, A.W. (1992). Univariate Discrete Distributions, 2nd Edition. John Wiley and Sons, New York

Ong, S.H. (1995). Computation of probabilities of a generalized log-series and related distributions. Commun. Statist. Theor. Meth., 24, 253-271.