Aligarh Journal of Statistics
Vol. 26 (2006), 9-18

ON A QUASI- NEGATIVE BINOMIAL DISTRIBUTION

A. Mishra and Anwar Hassan

ABSTRACT

A quasi-negative binomial distribution, in which the probability of an event is linearly dependent on the number of successes, has been studied. It has been found as a two-parameter gamma mixture of the 'generalized Poisson distribution' of Consul and Jain (1973).Some other models leading to this distribution have been given. Its moments have been obtained in terms of factorial power series. The estimation of its parameters has been discussed and the distribution has been fitted to some observed sets of data to test its goodness of fit.

1. INTRODUCTION

In a classical probability model an event is regarded as the result of a pure chance mechanism and the probability of a success at each trial is assumed to be constant. But this assumption does not appear to be realistic for many practical purposes. It is generally observed that living beings use their past experience and wisdom in determining strategies to increase or decrease their efforts for success or failure towards their goals and hence the probability of a success does not remain constant.

During the past twenty-five years or so, much work seems to have been done on the discrete probability models in which the probability of success of an event is linearly dependent on the number of successes. Among these, the quasi-binomial distribution (QBD) of Consul (1974), given by the probability function

$$P_1(x, p_1; p_2) = \binom{n}{x} p_1 (p_1 + x\, p_2)^{x-1} (1 - p_1 - x\, p_2)^{n-x}, \qquad x = 0, 1, 2, \cdots, n$$

$$0 < p_1 < 1\,, \; -p_1 / n < p_2 < (1 - p_1)/n \qquad (1.1)$$

and the generalized Poisson distribution (*GPD*) of Consul and Jain (1973), given by the probability function

$$P_2(x; \lambda_1; \lambda_{21}) = \lambda_1 (\lambda_1 + x\,\lambda_2)^{x-1} e^{-(\lambda_1 + x\lambda_2)} / x!,$$

$$x = 0, 1, 2, \cdots, \;\; \lambda_1 > 0\,, \; |\lambda_2| < 1 \qquad (1.2)$$

are the models which have received much attention of the research workers in Statistics. These distributions have been found to possess tremendous capability of explaining observed sets of discrete data from various fields. [Consul and Jain (1973), Mishra and Sinha (1981), Mishra *et al.* (1992)]

In response to the Nelson's (1975) comment that in the *GPD* (1.2) for $\lambda_2 < 0$ the probabilities sooner or later become negative, Consul and Shoukri (1985) recommended the use of truncation when $\lambda_2 < 0$ and the support of the distribution was restricted to $1, 2, \cdots, m$ where m is the largest integer for which $\lambda_1 + m\lambda_2 > 0$.

A negative binomial analogue of the *QBD* and *GPD* known as the quasinegative binomial distribution (*QNBD*), the probability function of which is given by

$$P_3(x; m; \alpha_1; \alpha_2) = \binom{m+x-1}{x} \alpha_1 (\alpha_1 + x\alpha_2)^{x-1} / (1 + \alpha_1 + x\alpha_2)^{m+x}$$

$$x = 0, 1, 2, \cdots, \quad \alpha_1 > 0, \quad |\alpha_2| < 1$$

such that

$$P_3(x; m; \alpha_1; \alpha_2) = 0 \text{ for } x \geq n \text{ if } \alpha_1 + n\alpha_2 < 0 \quad (1.3)$$

has though been mentioned by Janardan (1975) and Nandi and Das (1994) , has not been studied in detail, so far. In this paper an attempt has been made to study some of the aspects of this distribution. The *QNBD* has been found as a two-parameter gamma mixture of the 'generalized Poisson distribution' of Consul and Jain (1973). An urn-model leading to this distribution has been given. The moments of the distribution have been obtained in terms of the factorial power-series defined by Mishra and Singh (1999). The maximum likelihood method of estimation and other techniques of estimation of its parameters have been discussed and the distribution has been fitted to some observed sets of data to test its goodness of fit.

2. MODELS LEADING TO QUASI- NEGATIVE BINOMIAL DISTRIBUTION

Gamma mixture of the *GPD*

Taking $\lambda_1 = \alpha$ and $\lambda_2 / \lambda_1 = \theta$, the *GPD* (1.2) can be put in its restricted form as

$$\alpha^x (1 + x\theta)^{x-1} e^{-\alpha(1+x\theta)} / x! \quad (2.1)$$

Suppose that α is varying according to the gamma distribution with probability density function

$$f(\alpha)=\frac{b^m}{\Gamma(m)}e^{-b\alpha}\,\alpha^{m-1}, \quad \alpha>0,\ b>0.$$

Then the gamma mixture of (2.1) is obtained as

$$P_4(x;m,b,\theta)=b^m\binom{m+x-1}{x}(1+x\theta)^{x-1}/(1+b+x\theta)^{m+x} \tag{2.2}$$

Taking $1/b=\alpha_1$ and $\theta/b=\alpha_2$ the *QNBD* (1.3) is obtained.

Urn-model

Let there be two urns, A and B. Let urn A contain $'a'$ white balls and B contain $'a'$ white and $'b'$ red balls. Further, let $'m'$ and $'\theta'$ be two given positive numbers. A player decides his strategy by choosing a positive integer k and adds $k\theta$ red balls to urn A, $k\theta$ white balls and $m\theta$ red balls to urn B. A player then draws once ball from urn A. If it is red then he is not allowed to draw any ball from urn B. If it is white, he is asked to draw balls from urn B with replacement until m red balls are drawn. The player is declared as the winner of the game if he gets m red balls in exactly $(m+k)$ draws.

The probability of winning the game is given by

$$P_5(k;m,a,b,\theta)=\binom{m+k-1}{k}a\,(a+k\theta)^{k-1}(b+m\theta)^m/(a+b+(m+k)\theta)^{m+k} \tag{2.3}$$

Substituting α_1 for $a/(b+m\theta)$ and considering k as a particular value of x, the *QNBD* (1.3) is obtained.

Member of the family of Abel's series distributions

A family of Abel's series distributions has been defined by Nandi and Das (1994) by the following representation for real valued parameters $'a'$ and $'b'$

$$P_6(x;a,b)=a\,(a-xb)^{x-1}\,\beta(x,b)/g(a) \tag{2.4}$$

where $g(a)$ is a finite and positive function given by

$$g(a)=\sum_{x=0}a\,(a-xb)^{x-1}\,\beta(x,b) \quad\text{and}\quad \beta(x,b)=\frac{1}{x!}\frac{d^x g(a)}{d a^x}\bigg|_{a=xb}$$

Selecting $g(a)=(c-a)^{-m}$ we get,

$$(c-a)^{-m}=\sum\binom{m+x-1}{x}a\,(a-xb)^{x-1}(c-xb)^{-(m+x)}$$

Taking $a/(c-a)=\alpha_1$ and $-b/(c-a)=\alpha_2$ the *QNBD* (1.3) is obtained.

Quasi-binomial distribution

It can be easily seen that the *QNBD* (1.3) can be put in the form

$$P_3(x;m,\alpha_1,\alpha_2)=\binom{-m}{x}(-\alpha_1)(-\alpha_1-x\alpha_2)^{x-1}(1+\alpha_1+x\alpha_2)^{-m-x} \quad (2.5)$$

which is the *QBD* (1.1) in which the parameters n, p_1 and p_2 have been replaced by $-m$, $-\alpha_1$ and $-\alpha_2$ respectively.

3. MOMENTS

The $r-$th moment about origin of the *QNBD* can be obtained in two stages as

$$M_r' = E(X^r) = E[E(X^r \mid \alpha)]$$

where $E(X^r \mid \alpha)$ is the conditional expectation given α.

The first two moments of the restricted form of the *GPD* (2.1) as obtained by Consul and Jain (1973) are

$$\mu_1' = \frac{\alpha}{(1-\alpha\theta)}; \quad \mu_2' = \frac{\alpha}{(1-\alpha\theta)^3}+\frac{\alpha^2}{(1-\alpha\theta)^2} \quad (3.1)$$

Using this, we find the mean of the *QNBD* as

$$M_1' = E\left(\frac{\alpha}{(1-\alpha\theta)}\right)=\frac{b^m}{\Gamma(m)}\int_0^\infty (1-\alpha\theta)^{-1}e^{-b\alpha}\alpha^m\, d\alpha$$

$$=\frac{b^m}{\Gamma(m)}\int_0^\infty \sum_{i=0}\theta^i e^{-b\alpha}\alpha^{m+i}d\alpha = \frac{b^m}{\Gamma(m)}\sum_{i=0}\theta^i\left(\int_0^\infty e^{-b\alpha}\alpha^{m+i}d\alpha\right)$$

$$=\frac{b^m}{\Gamma(m)}\sum_{i=0}\theta^i\,\frac{\Gamma(m+i+1)}{b^{m+i+1}}=\frac{m}{b}\sum_{i=0}(m+1)^{(-i)}\left(\frac{\theta}{b}\right)^i$$

$$=m\alpha_1\left(\sum_{i=0}(m+1)^{(-i)}\alpha_2^i\right)$$

where as earlier $\alpha_1 = 1/b$, $\alpha_2=\theta/b$ and $a^{(-i)}$ is defined as

$$a^{(-j)} = a(a+1)(a+2)\cdots(a+j-1)$$

Factorial power series

For finding the moments of the *QBD*, Mishra and Singh (1999) defined a series termed as ‘factorial power series’, $J(s,a,b)$, of order 's' in 'a' and 'b' as

$$J(s,a,b) = \sum C_j(s)\, a^{(j)} b^j \tag{3.2}$$

where $'s'$ is a positive integer , $C_j(s)$ is the coefficient of y^j in the expansion of $(1-y)^{-s}$ and $a^{(j)} = a(a-1)(a-2)(a-3)\cdots(a-j+1)$

We define here a factorial power series, $J'(s,a,b)$, of order $'s'$ in $'a'$ and $'b'$ in a slightly different way as

$$J'(s,a,b) = \sum C_j(s)\, a^{(-j)} b^j \tag{3.3}$$

where $'s'$ and $C_j(s)$ have the same meaning.

Using this factorial power series, the mean of the *QNBD* can be expressed as

$$M_1' = m\alpha_1\, J'(1, m+1, \alpha_2) \tag{3.4}$$

The second moment about origin of the *QNBD* is similarly obtained as

$$M_2' = E\left(\frac{\alpha}{(1-\alpha\theta)^3} + \frac{\alpha^2}{(1-\alpha\theta)^2}\right)$$

$$= \frac{b^m}{\Gamma(m)}\left(\int_0^\infty (1-\alpha\theta)^{-3} e^{-b\alpha} \alpha^m \, d\alpha + \int_0^\infty (1-\alpha\theta)^{-2} e^{-b\alpha} \alpha^{m+1}\, d\alpha\right)$$

$$= \frac{m}{b}\sum_{i=0}\binom{-3}{i}(m+1)^{(-i)}\left(\frac{-\theta}{b}\right)^i + \frac{m(m+1)}{b^2}\sum_{i=0}\binom{-2}{i}(m+2)^{(-i)}\left(\frac{-\theta}{b}\right)^i$$

$$= m\,\alpha_1 \sum_{i=0}\binom{-3}{i}(m+1)^{(-i)}(-\alpha_2)^i + m(m+1)\,\alpha_1^2 \sum_{i=0}\binom{-2}{i}(m+2)^{(-i)}(-\alpha_2)^i$$

$$= m\alpha_1\, J'(3, m+1, \alpha_2) + m(m+1)\,\alpha_1^2\, J'(2, m+2, \alpha_2) \tag{3.5}$$

Mishra and Singh (1999) obtained the first four moments of the *QBD* (1.1) in terms of factorial power series. In the moments of the *QBD* if n, p_1 and p_2 are replaced by $-m$, $-\alpha_1$ and $-\alpha_2$ as in (2.5) the moments of the *QNBD* can be obtained. It can be verified that the same expressions for the first two moments obtained in (3.4) and (3.5) are obtained from the first two moments of the *QBD* after such substitutions. Similarly, the third and the fourth moments of the *QNBD* are obtained as

$$M_3' = m\alpha_1 [J'(4, m+1, \alpha_2) + 3(m+1)\alpha_2\, J'(5, m+2, \alpha_2)]$$

$$+ 3\,m^{(-2)}\,\alpha_1^2\, J'(4, m+2, \alpha_2) + m^{(-3)}\,\alpha_1^3\, J'(3, m+3, \alpha_2) \tag{3.6}$$

$$M'_4 = m\alpha_1[J'(7, m+1, \alpha_2) + 8(m+1)\,\alpha_2\, J'(7, m+2, \alpha_2)]$$
$$+ m^{(-2)}\,\alpha_1^2\,[7\,J'(6, m+2, \alpha_2) + 8(m+2)\,\alpha_2\, J'(6, m+3, \alpha_2)]$$
$$+ 6m^{(-3)}\,\alpha_1^3\, J'(5, m+3, \alpha_2) + m^{(-4)}\,\alpha_1^4\, J'(4, m+4, \alpha_2)$$
$$+ 6\,(m+1)(m+2)\,\alpha_2^2\, J'(7, m+3, \alpha_2) \quad (3.7)$$

4. ESTIMATION OF PARAMETERS

i) Using the first three frequencies

Let $P(0)$, $P(1)$ and $P(2)$ denote the probabilities for $x=0$, $x=1$ and $x=2$ respectively according to the *QNBD* (1.3). We find that

$$\alpha_1 = [P(0)]^{-1/m} - 1 \quad (4.1)$$

$$\alpha_2 = [m\alpha_1 / P(1)]^{1/(m+1)} - \alpha_1 - 1 \quad (4.2)$$

$$m = 2\,P(2)(1+\alpha_1+2\alpha_2)^{m+2} / (m+1)\,\alpha_1\,(\alpha_1 + 2\alpha_2) \quad (4.3)$$

We can find a function such that

$$\phi(m) = 2\;P(2)(1+\alpha_1+2\alpha_2)^{m+2} / (m+1)\,\alpha_1\,(\alpha_1 + 2\alpha_2) - m = 0 \quad (4.4)$$

Replacing $P(0)$, $P(1)$ and $P(2)$ by the corresponding observed relative frequencies $p(0)$, $p(1)$ and $p(2)$ respectively and taking an initial value of m in (4.1), a value of α_1 can be obtained which and the value of m when substituted in (4.2) , a value of α_2 can be obtained. Substituting these values of m, α_1 and α_2 in (4.4) a value of $\phi(m)$ can be obtained. The equation (4.4) can be solved for the value of m by using any iterative method such as regula-falsi method. Then for this value of m, the values of α_1 and α_2 can obtained. It may be mentioned here that if any frequency of the first three classes is zero, this method fails to provide the estimates of the parameters.

ii) Maximum likelihood estimation

The log-likelihood function of the *QNBD* is obtained as

$$Log\ L = \sum_{x=0}^{n}\sum_{j=1}^{x} f_x \log(m+x-j) + N\log\alpha_1 + \sum_{x=0}^{n} f_x (x-1)\log(\alpha_1 + x\alpha_2)$$
$$- \sum_{x=0}^{n}(m+x)\,f_x \log(1+\alpha_1+x\alpha_2) - \sum_{x=0}^{n} f_x \log x! \quad (4.5)$$

and the three likelihood equations as

$$\frac{\partial Log\ L}{\partial m} = \sum_{x=0}^{n}\sum_{j=1}^{x} \frac{f_x}{(m+x-j)} - \sum_{x=0}^{n} f_x \log(1+\alpha_1 + x\alpha_2) = 0 \tag{4.6}$$

$$\frac{\partial Log\ L}{\partial \alpha_1} = \frac{N}{\alpha_1} + \sum_{x=0}^{n}(x-1)\frac{f_x}{(\alpha_1 + x\alpha_2)} - \sum_{x=0}^{n}\frac{(m+x) f_x}{(1+\alpha_1 + x\alpha_2)} = 0 \tag{4.7}$$

$$\frac{\partial Log\ L}{\partial \alpha_2} = \sum_{x=0}^{n} x(x-1)\frac{f_x}{(\alpha_1 + x\alpha_2)} - \sum_{x=0}^{n} x(m+x)\frac{f_x}{(1+\alpha_1 + x\alpha_2)} = 0 \tag{4.8}$$

where f_x is the frequency of x.

These equations are not simple to provide direct solution and thus can be solved using the Fisher's scoring method .For; we can solve the following system of equations

$$\left[\left(\frac{-\partial^2\ Log\ L}{\partial \theta^2}\right)\right]_{\theta_0} (\hat{\theta} - \theta_0) = \left[\left(\frac{-\partial\ Log\ L}{\partial \theta}\right)\right]_{\theta_0} \tag{4.9}$$

where $\theta' = (m,\ \alpha_1, \alpha_2)$ is a parameter vector, $\hat{\theta}$ is the vector of ML estimates of θ and θ_0 is a vector of trial values of θ. The trial values of the parameters may be obtained on the basis of the first three class frequencies.

Multiplying (4.7) by α_1 and (4.8) by α_2 and then adding we find a nice result as

$$m = \frac{1}{N}\sum_{x=0}^{n}\frac{(m+x)}{(1+\alpha_1 + x\alpha_2)} f_x \tag{4.10}$$

and substituting it in (4.7) we find

$$m = \frac{1}{\alpha_1} + \frac{1}{N}\sum_{x=1}^{n}\frac{(x-1)}{(\alpha_1 + x\alpha_2)} f_x \tag{4.11}$$

iii) Using the first two frequencies and a likelihood equation

We have from (4.11)

$$g(m) = \frac{1}{\alpha_1} + \frac{1}{N}\sum_{x=1}^{n}\frac{(x-1)}{(\alpha_1 + x\alpha_2)} f - m = 0 \tag{4.12}$$

This also can be solved using the regula-falsi method. Taking an initial value of m in (4.1) and so obtaining a value of α_1, a value of α_2 can be obtained from (4.2) which when substituted in (4.12) a value of $g(m)$ is obtained.

5. GOODNESS OF FIT

As in the *QNBD* the probability of a success does not remain constant and depends upon the previous number of successes, the *QNBD* seems to be more realistic than the classical negative binomial distribution and so is expected to explain data in a better way than the negative binomial distribution.

The *QNBD* has been fitted to a number of observed data-sets and in general it was found to give better fits than the negative binomial distribution. We present here only two of such cases. The estimates of the parameters have been obtained using the first two frequencies and a likelihood equation as discussed in (iii) of the previous section. In each of the following two tables the expected frequencies according to the negative binomial distribution are also given so that a quick comparison can be made.

It may be mentioned here that Jain and Consul (1971) fitted their generalized negative binomial distribution (*GNBD*) to the same data-sets as given in tables-1 and 2 to show the superiority of the *GNBD* over the logarithmic series distribution. The values of χ^2 for these data-sets in case of the *GNBD* were reported as 2.35 and 2.07 whereas these values in the case of the *QNBD* are 1.94 and 1.99 respectively at the same *df* .

Table 1: Accidents to 647 women working on H.E. Shells during 5 weeks

Number of accidents	Observed frequency	Expected frequency	
		NBD	*QNBD*
0	447	442.9	447.0
1	132	138.6	132.0
2	42	44.4	44.8
3	21	14.3	16.4
4	3	4.6	4.7
5	2	2.2	2.1
Total	647	647.0	647.0
χ^2		4.09	1.94
df		2	1

$M_1' = 0.465224$, $M_2 = 0.691900$, $\hat{m} = 0.672572$, $\hat{\alpha}_1 = 0.732926$, $\hat{\alpha}_2 = -0.038331$.

Table 2: Counts of the number of European red mites on apple leaves

Number of Mites per leaf	Leaves Observed	Expected frequency	
		NBD	*QNBD*
0	70	67.5	70.0
1	38	39.0	38.0
2	17	21.0	19.6
3	10	11.0	10.1
4	9	5.6	5.9
5	3	2.9	3.2
6	2	1.5	1.8
7	1	0.8	0.9
8	0	0.7	0.5
Total	150	150.0	150.0
χ^2		2.93	1.99
df		3	2

$M_1' = 1.146667$, $M_2 = 2.273646$, $\hat{m} = 1.236818$, $\hat{\alpha}_1 = 0.851898$, $\hat{\alpha}_2 = 0.039284$.

REFERENCES

Consul, P.C. (1974): A simple urn-model dependent upon pre-determined strategy. *Sankhyā, Ser B*, **36**, 391-399.

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

Consul, P.C. and Shoukri, M.M. (1985): The generalized Poisson distribution when the sample mean is larger than the sample variance. *Comm. Statist. Simulation Comput*., **14**, 667-681.

Jain, G.C. and Consul, P.C. (1971): A generalized negative binomial distribution, *SIAM J. Appl. Math.,* **21**, 501-512.

Janardan, K.G. (1975): Markov-Polya urn–model with pre-determined strategies. *Gujarat Statist. Rev.*, **2**, 17-32.

Mishra, A and Sinha, J.K. (1981): A generalization of binomial distribution. *J. Indian Statist. Assoc*., **19**, 93-98.

Mishra, A. and Singh, S.K. (1999): Moments of a quasi-binomial distribution. *Assam Statist. Rev.*, **13**, 13-20.

Mishra, A; Tiwary, D. and Singh, S.K. (1992): A class of quasi-binomial distribution. *Sankhyā, Ser B*, **54**, 67-76.

Nandi, S.B. and Das, K.K. (1994): A family of the Abel series distributions. *Sankhyā, Ser B*, **56**, 147-164.

Nelson, D.L. (1975): Some remarks on generalization of the negative binomial and Poisson distributions. *Technometrics*, **17,** 135-136.

Received : 30-01-2003

Revised : 21-07-2006

A. Mishra
Patna University, Patna-800 005 , India
e-mail: mishraamar@rediffmail.com

Anwar Hassan
P.G. Department of Statistics
University of Kashmir, Srinagar, India