Vol.6.Issue.3.2018 (July-Sept.)

BULLETIN OF MATHEMATICS AND STATISTICS RESEARCH

A Peer Reviewed International Research Journal

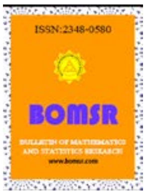
http://www.bomsr.com
Email:editorbomsr@gmail.com

INTERNATIONAL STANDARD SERIAL NUMBER 2348-0580

RESEARCH ARTICLE

A GENERALISED NEGATIVE BINOMIAL DISTRIBUTIONAND ITS IMPORTANT FEATURES

BINOD KUMAR SAH
Reader in Statistics, R.R.M. Campus, Janakpur,Tribhuvan University, Nepal
Email- sah.binod01@gmail.com



ABSTRACT
In this paper, generalization of Negative binomial distribution has been obtained by mixing generalized Poisson distribution of Consul and Jain (1973) with the two-parameter gamma distribution. Its probability density function has been obtained. The first four moments of the distribution have been obtained and the estimation of its parameters by the method of moments has been discussed. Important features of this distribution have been discussed.This distribution has been found more general in nature and wider in scope than the negative binomial distribution. The distribution has been fitted to some well known data-sets having variance greater than mean and it has been observed that the generalized negative binomial distribution gives better fit than the negative binomial distribution.

Keywords: generalized Poisson distribution, compounding, negative binomial distribution, goodness of fit, moments

1. INTRODUCTION

In the theory of Statistics, gamma distribution plays a very important role. It is a member of two- parameter family of continuous probability distributions. The probability density function of gamma distribution can be expressed in terms of the gamma function parameterized in terms of a shape parameter 'm' and scale parameter 'a'. A continuous random variable x is said to follow two – parameter gamma distribution with parameter m and a, if its probability density function is given by

$$g\left(x;a,m\right)=\frac{a^{m}}{\Gamma\left(m\right)}.x^{m-1}\,e^{-ax} \tag{1.1}$$

For $x \geq 0$ and m, a > 0

For a = 1, it reduces to the one – parameter gamma distribution and for m = 1 it reduces to the exponential distribution. The r^{th} moment about origin of gamma distribution is obtained as

$\mu_r^{/} = m^{(-r)} / a^r$; Where $m^{(-r)}$ = m (m+1) ... (m+r+1) (1.2)

and the mean and variance of the gamma distribution are m/a and m/a^2 respectively.

Gamma distribution has also been used as a mixing distribution with the various probability distributions. It is well known that a continuous mixture of Poisson distribution where the mixing distribution is a gamma distribution results in a negative binomial distribution. That is, we can view the negative binomial as a Poisson (λ) distribution, where λ is itself a random variable, distributed according to Gamma (a, m) which can symbolically be shown as

NBD (m, p) = Poisson (λ) $\underset{\lambda}{\wedge}$ Gamma (a, m) (1.3)

Where p = $\dfrac{a}{(1+a)}$

Negative binomial distribution has been found to provide useful representation in accident Statistics, birth and death process. There are of course situations where a good fit is not obtainable with negative binomial distribution, and in such cases it is usual to consider the possibility of a mixture of distributions or a contagious distribution having more than two parameters. In the present paper, a generalization of negative binomial distribution has been obtained by mixing the GPD with the two – parameter gamma distribution. The various aspects of the resultant distribution such as its moments, estimation of parameters, goodness of fit, have also been obtained.

2 GENERALISED NEGATIVE BINOIAL DISTRIBUTION (GNBD)

A generalisation of the negative binomial distribution can be obtained by taking the Consul and Jain's GPD in place of the classical Poisson distribution which can symbolically be shown as

GPD (λ,θ) $\underset{\lambda}{\wedge}$ Gamma (a , m) (2.1)

The probability density function of this Gamma mixture of GPD is obtained as follows.

$$P(x) = \int_0^\infty \frac{\lambda(\lambda+\theta x)^{x-1} e^{-(\lambda+\theta x)}}{x!} . \frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)} d\lambda \quad (2.2)$$

$$= \frac{a^m e^{-\theta x}}{\Gamma(m) x!} \int_0^\infty e^{-(1+a)\lambda} \lambda^{m+x-1} \left(1+\frac{\theta x}{\lambda}\right)^{x-1} d\lambda$$

$$= \frac{\left(\frac{a}{1+a}\right)^m \left(\frac{1}{1+a}\right)^x e^{-\theta x}}{\Gamma(m)\cdot x!} \cdot \sum_{i=0}^{x-1} \binom{x-1}{i} (\theta x)^i \frac{\Gamma(m+x-i)}{(1+a)^{-i}}$$

Let a/(1+a) = p and 1/(1+a) = q

$$= e^{-\theta x} p^m \sum_{i=0}^{x-1} \frac{(x-1)!}{i!(x-i-1)!} \cdot \frac{(m+x-i-1)!}{(m-1)!\, x!} (\theta x)^i \, q^{x-i} \quad 2.2)$$

$$= e^{-\theta x} p^m \binom{m+x-1}{x} q^x + \theta p^m e^{-\theta x} \sum_{i=1}^{x-1} \frac{(m+x-i-1)!}{(x-i-1)!(m-1)!} \frac{(\theta x)^{i-1}}{i!} q^{x-i} \quad (2.3)$$

The expression (2.3) is the probability density function of a gamma mixture of GPD. At $\theta = 0$, it reduces to probability mass function of negative binomial distribution. Hence, the relation (2.3) may be called as a generalised negative binomial distribution (GNBD). It can also be expressed as

$$P(x) = p^m \; for \; x = 0$$

$$= \frac{mp^m e^{-\theta x}}{x} \sum_{i=1}^{x-1} \binom{m+x-i-1}{x-i-1} \frac{q^{x-i} (\theta x)^{x-i}}{\Gamma(i+1)} \quad \text{for x =1,2,3,} \quad (2.4)$$

Putting m=1 in the relation (2.4), we get probability density function of a new generalised form of the geometric distribution given by Mishra (2009) as

$$P_1(x)=\frac{pe^{-\theta x}}{x}\sum_{i=1}^{x-1}\frac{(x-i)(\theta x)^i}{\Gamma(i+1)}\cdot q^{x-i} \quad \text{for } x = 1,2,3, \tag{2.5}$$

3. MOMENTS OF GNBD

The r^th moment about origin of the generalised form (2.3) is obtained as

$$\mu_r^{/}=E\left[E\left(x^r/\lambda\right)\right]$$

$$\mu_r^{/}=\int_0^{\infty}\left[\sum_{x=0}^{\infty}\frac{x^r\lambda(\lambda+\theta x)^{x-1}e^{-(\lambda+\theta x)}}{\Gamma(x+1)}\right]\frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)}\,d\lambda \tag{3.1}$$

Obviously the expression under bracket is the r^th moment about origin of the GPD. Taking r = 1 and using the mean of GPD for the expression in bracket, the first moment about origin of the GNBD is obtained as

$$\mu_1^{/}=\int_0^{\infty}\frac{\lambda}{(1-\theta)}\frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)}d\lambda \tag{3.2}$$

$$\text{i.e } Mean\ (\mu_1^{/})=\frac{m}{(1-\theta)a}=\frac{mq}{(1-\theta)p} \tag{3.3}$$

Taking r = 2 in (3.1) and using the second moment about origin of the GDP, the second moment about origin of the GNBD can be obtained as

$$\mu_2^{/}=\int_0^{\infty}\left[\left\{\frac{\lambda}{(1-\theta)^3}+\frac{\lambda^2}{(1-\theta)^2}\right\}\left\{\frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)}\right\}\right]d\lambda \tag{3.4}$$

$$=\frac{m}{(1-\theta)^3 a}+\frac{(m+1)m}{(1-\theta)^2 a^2}=\frac{mq}{(1-\theta)^3 p}+\frac{(m+1)m}{(1-\theta)^2}\cdot\frac{q^2}{p^2} \tag{3.5}$$

Taking r = 3 in (3.1) and using the third moment about origin of the GPD for the expression in bracket, the third moment about origin of the GNBD can be obtained as

$$\mu_3^{/}=\int_0^{\infty}\left\{\frac{\lambda(1+2\theta)}{(1-\theta)^5}+\frac{3\lambda^2}{(1-\theta)^4}+\frac{\lambda^3}{(1-\theta)^3}\right\}\frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)}\,d\lambda \tag{3.6}$$

$$=\frac{(1+2\theta)m}{(1-\theta)^5}\cdot\frac{q}{p}+\frac{3(m+1)m}{(1-\theta)^4}\cdot\frac{q^2}{p^2}+\frac{(m+2)(m+1)m}{(1-\theta)^3}\cdot\frac{q^3}{p^3} \tag{3.7}$$

Similarly, taking r = 4 in (3.1) and using the fourth moment about origin of the GPD for the expression in bracket, the fourth moment about origin of the GNBD can be obtained as

$$\mu_4^{/}=\int_0^{\infty}\left[\frac{\lambda(1+8\theta+6\theta^2)}{(1-\theta)^7}+\frac{\lambda^2(7+8\theta)}{(1-\theta)^6}+\frac{6\lambda^3}{(1-\theta)^5}+\frac{\lambda^4}{(1-\theta)^4}\right]\cdot\frac{a^m e^{-a\lambda}\lambda^{m-1}}{\Gamma(m)}\,d\lambda \tag{3.8}$$

$$\begin{aligned}&=\frac{m(1+8\theta+6\theta^2)}{(1-\theta)^7}\cdot\frac{q}{p}+\frac{m(m+1)(7+8\theta)}{(1-\theta)^6}\cdot\frac{q^2}{p^2}\\&+\frac{6m(m+1)(m+2)}{(1-\theta)^5}\cdot\frac{q^3}{p^3}+\frac{m(m+1)(m+2)(m+3)}{(1-\theta)^4}\cdot\frac{q^4}{p^4}\end{aligned} \tag{3.9}$$

The central moments of the GNBD has thus been obtained as

$$\mu_2 = \mu_2' - \mu_1'^2$$

$$= \frac{m}{(1-\theta)^3}\cdot\frac{q}{p} + \frac{(m+1)m}{(1-\theta)^2}\cdot\frac{q^2}{p^2} - \left[\frac{m}{1-\theta}.\frac{q}{p}\right]^2$$

$$= \frac{m}{(1-\theta)^3}\cdot\frac{q}{p} + \frac{m}{(1-\theta)^2}\cdot\frac{q^2}{p^2} \qquad (3.10)$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 3\mu_1'^3$$

$$= \frac{(1+2\theta)m}{(1-\theta)^5}\cdot\frac{q}{p} + \frac{3(m+1)m}{(1-\theta)^4}\cdot\frac{q^2}{p^2} + \frac{(m+2)(m+1)m\,q^3}{(1-\theta)^3\,p^3}$$

$$-3\left[\frac{m}{(1-\theta)^3}\cdot\frac{q}{p} + \frac{(m+1)m}{(1-\theta)^2}\cdot\frac{q^2}{p^2}\right]\left[\frac{m}{(1-\theta)}\cdot\frac{q}{p}\right] + 2\left[\frac{m}{(1-\theta)}\cdot\frac{q}{p}\right]^3$$

on a little simplification , we get

$$\mu_3 = \frac{(1+2\theta)m}{(1-\theta)^5}\cdot\frac{q}{b} + \frac{3m}{(1-\theta)^4}\cdot\frac{q^2}{p^2} + \frac{2m}{(1-\theta)^3}\cdot\frac{q^3}{p^3} \qquad (3.11)$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

After a little simplification, we get

$$\mu_4 = \frac{m(1+8\theta+6\theta^2)}{(1-\theta)^7}\cdot\frac{q}{p} + \frac{m(3m+7+8\theta)}{(1-\theta)^6}\cdot\frac{q^2}{p^2}$$

$$+\frac{m(6m+12)}{(1-\theta)^5}\cdot\frac{q^3}{p^3} + \frac{m(6+3m)}{(1-\theta)^4}\cdot\frac{q^4}{p^4} \qquad (3.12)$$

Important features of the GNBD

(a) From mean and variance of GNBD

$$(1-\theta)^2\,\mu_2 = \mu_1' + \frac{mq^2}{p^2}$$

$$i.e.(1-\theta)^2\,\mu_2 - \mu_1' = \frac{mq^2}{p^2} > 0$$

This gives the inequality

$$\mu_2 > \mu_1' / (1-\theta)^2 \qquad (3.13)$$

The parameter θ appears to play an important role in explaining the inequality between mean and variance. At θ=0, we have $\mu_2 > \mu_1'$ which is the characteristic of the negative binomial distribution. But for non-zero values of θ the lower limit for the variance becomes different from mean. For $0<\theta<1$, the lower limit of the variance is higher than the mean, obviously providing an improved lower limit of the variance. Similarly for $-1<\theta<0$ the lower limit of the variance is lower than the mean. Thus the parameter θ has capacity to stretch the lower limit of the variance up and down according to the observed distribution. Obviously for non-negative value of θ the inequality between mean and variance is explained in a better way than that explained by the negative binomial distribution. Thus the parameter θ takes into account the inequality between mean and variance more closely than that taken by NBD and so it is expected that in all such cases where variance is greater than mean , the GNBD would explain observed data more closely than the negative binomial distribution.

(b) Generalised Logarithmic Series Distribution (GLSD) obtained by Binod Kumar Sah(2013) is the limiting form of zero-truncated GNBD:

Proof:

Probability mass function of the zero-truncated GNBD can be obtained from (2.3) as

$$P_2(x) = \frac{P(x)}{1-p^m}$$

$$= \frac{mp^m e^{-\theta x}}{(1-p^m)x} \cdot \sum_{i=0}^{x-1} \binom{m+x-i-1}{x-i-1} \frac{q^{x-i}(x\theta)^i}{i!} \qquad (3.14)$$

This can be put in expanded form as x = 1 , 2 , 3 , ...

$$P_2(x) = \frac{mp^m e^{-\theta x}}{(1-p^m)x} \cdot \sum_{i=0}^{x-1} (m+x-i-1)(m+x-i-2)\ldots(m+1)m(m-1)!$$

$$\frac{q^x(\theta x)^i}{(x-i-1)m(m-1)!\,i!}$$

Taking limit $m \to 0$, we get

$$\underset{m\to 0}{Lt} P_2(x) = \frac{e^{-\theta x}}{x} \cdot \underset{m\to 0}{Lt} \frac{m}{(1-p^m)} \cdot \underset{m\to 0}{Lt} p^m \cdot \underset{m\to 0}{Lt}$$

$$\sum_{i=0}^{x-1} \frac{(m+x-i-1)(m+x-i-2)\cdots(m+1)q^{x-i}(\theta x)^i}{(x-i-1)!\,i!}$$

i.e $$\underset{m\to 0}{Lt} P_2(x) = -\frac{1}{\log(1-q)} \cdot \frac{q^x e^{-\theta x}}{x} \cdot \sum_{i=0}^{x-1} \frac{(\theta x)^i q^{-i}}{i!} \qquad (3.15)$$

It is the limiting form of zero-truncated GNBD. It reduces to the Logarithmic series distribution at θ =0, we term it as a generalized logarithmic series distribution (GLSD).

4. ESTIMATION OF PARAMETERS

The generalized mixture distribution (2.3) consists of three parameters m, p and θ and so for estimating these parameters by the method of moments, the first three moments are required.

The variance (μ_2) of the GNBD can also be expressed as

$$\mu_2 = \frac{mq}{(1-\theta)p}\left[\frac{1}{(1-\theta)^2} + \frac{q}{(1-\theta)p}\right] \qquad (4.1)$$

Let,s = $(1-\theta)^{-1}$ in (4.1) , we get

$$\frac{\mu_2}{\mu_1^{/}} = s^2 + sa = K_1\,(Say) \qquad (4.2)$$

i.e. a = $\frac{K_1 - s^2}{s}$ (4.3)

Using the third central moment and the mean of the GNBD, we have

$$\frac{\mu_3}{\mu_1^{/}} = \left(1+2\theta\right)s^4 + 3s^3 a + 2s^2 a^2 = K_2\,(Say) \quad (4.4)$$

Substituting the value of $\theta = \frac{s-1}{s}$ and a = $\frac{K_1 - s^2}{s}$ in (4.4) and after a little simplification, we get

$$f\left(s\right) = 2s^4 - 2s^3 - K_1 s^2 + \left(2K_1^2 - K_2\right) = 0 \quad (4.5)$$

This is a polynomial in s of the fourth degree and may be solved by using some iterative method such as the Newton – Raphson method or the Regula – Falsi method.

The estimates of the parameters m, p and θ can also be obtained by using P(x=0), $\mu_1^{/}$ and μ_2.

Taking logarithm both sides of the first relation of (2.4), we get

Log P(x=0) = m log p

$$\text{i.e. } m = \frac{\log P\left(x=0\right)}{\log p} \quad (4.6)$$

The mean of the GNBD can also be written as

$$\mu_1^{/} = \text{mas} \qquad \text{i.e. } m = \frac{\mu_1^{/}}{as} \quad (4.7)$$

Equating the relations (4.6) and (4.7) , we get

$$\frac{\mu_1^{/}}{as} = \frac{\log P\left(x=0\right)}{\log p} \qquad \text{i.e. } s = \frac{\mu_1^{/}\log p}{a\log P\left(x=0\right)} \quad (4.8)$$

Let $\frac{\log P\left(x=0\right)}{\mu_1^{/}} = K_3$. The relationcan be written as

$$s = \frac{\log p}{a\,K_3} \quad (4.9)$$

Also,

$$p = \frac{1}{\left(1+a\right)} = \frac{s}{K_1 - s^2 + s}\; and\; q = \frac{K_1 - s^2}{K_1 - s^2 + s} \quad (4.10)$$

Substituting the value of *p* and *a*in (4.9), we get an estimating equation for *s* as

$$s = \frac{\log\left(\frac{s}{K_1 - s^2 + s}\right)}{K_3\left(\frac{K_1 - s^2}{s}\right)} \quad . \qquad \text{Or,} \qquad sK_3\,(K_1 - s^2) = s\log\left(\frac{s}{K_1 - s^2 + s}\right)$$

$$\text{i.e. } f\left(s\right) = K_3\left(K_1 - s^2\right) - \log\left(\frac{s}{K_1 - s^2 + s}\right) = 0 \quad (4.11)$$

After replacing the population moments by the corresponding sample moments, this equation may be solved by the Newton –Raphson method or Regula –Falsi method for the value of s which gives an estimate of θ. Substituting this value of s in (4.10) the estimate of q and so of p can be obtained. Substituting these estimates of θ and p in the expression for mean, an estimate of m can be obtained.

5. GOODNESS OF FIT

The GNBD is supposed to have its applications to all those cases where the negative binomial distribution has. The GNBD was fitted to a number of data-sets where previously negative binomial distribution was used by others. It was found that the fits given by the GNBD were closer than those given by negative binomial distribution. Only three of them are being reported here.

In table 1, the data –set of Greenwood and Yule(1920) on the number of accidents to 647 women working on H.E. Shells during five weeks is considered. In table 2 , the Garman data on counts of European red mites on apple leaves and in table 3 , the Bortkewitch's data on the number of deaths caused by horse – kicks in the Prussian army have been considered.

In the first two data – sets the mean is less than the variance indicating the case of negative binomial distribution and in the third data – set as mean is not significantly different from the variance , so it is the case of Poisson distribution.

The observed and expected frequencies according to the GNBD have been given in the tables. For quick comparisons the expected frequencies according to the negative binomial distribution have also been given in the respective tables 1 and 2 and those according to the Poisson distribution in the table 3.

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

Number of accidents	Observed Frequency	Expected Frequency	
		Negative Binomial	GNBD
0	447	442.9	443.7
1	132	138.6	136.8
2	42	44.4	45.1
3	21	14.3	15.5
4	3	4.6	4.0
5+	2	2.2	1.9
Total	647	647.0	647.0
μ_1' = 0.4652241		χ^2 = 4.09	2.41
μ_2 = 0.6919002			
μ_3 = 1.2179804		d.f. = 3	2
$\hat{m}$ = 1.347201			
$\hat{p}$ = 0.755768		P(χ^2) = 0.25	0.30
$\hat{\theta}$ = 0.0641993			

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

Number of red mites / leaf	Observed Frequency	Expected Frequency	
		Negative Binomial	GNBD
0	70	67.5	68.2
1	38	39.0	37.9
2	17	21.0	20.8
3	10	11.0	11.2
4	9	5.7	6.7
5	3	2.8	2.7
6	2	1.5	1.7

7+	1	1.5	1.0
Total	150	150.0	150.0
μ_1' = 1.1466667		χ^2 = 2.93	1.72
μ_2 = 2.2584888			
μ_3 = 5.189777		d.f. = 4	3
$\hat{m}$ = 1.983821			
$\hat{p}$ = 0.672061		P(χ^2) = 0.58	0.65
$\hat{\theta}$ = 0.156211			

Table 3: Deaths due to horse-kicks in the Prussian Army

Number of deaths	Observed Frequency	Expected Frequency	
		Poisson	GNBD
0	109	108.7	110.4
1	65	66.3	63.7
2	22	20.2	21.1
3	3	4.1	4.2
4	1	0.7	0.6
Total	200	200.0	200.0
μ_1' = 0.6100000		χ^2 = 0.32	0.22
μ_2' = 0.6109548			
μ_3' = 0.5905620		d.f. = 3	2
$\hat{m}$ = 21.650121			
$\hat{p}$ = 0.972991		P(χ^2) = 0.96	0.90
$\hat{\theta}$ = 0.014542			

6. CONCLUSION

Comparing the expected frequencies with the observed ones and also by their chi-square values and p-values, it is evident that the GNBD provides very close fit to the data-sets. This distribution gives closer fit to the discrete data of negative binomial in nature than the classical negative binomial distribution. Generalised Logarithmic Series Distribution (GLSD) obtained by Binod Kumar Sah(2013) is the limiting form of zero-truncated GNBD.

Acknowledgement

The author expresses their gratitude to the referee for his valuable comments and suggestions which improved the quality of the research article. The hearty thank goes to Professor A. Mishra for his tremendous guidance and valuable suggestions.

REFERENCES

[1]. Consul, P.C., and Jain, G.C. (1973a): A generalisation of the Poisson distribution, Technometrics, 15, 791-799.

[2]. Greenwood, M., and Yule, G.U. (1920): An inquiry into the nature of frequency distribution representative of multiple happenings with particular reference to the occurrence of multiple attack sof disease or of repeated accidents, Journal of the Royal Statistical Society, Series A, 83, 255-279.

[3]. Mishra, A. (2009): A new generalized form of geometric distribution, International Journal of Agricultural and Statistical Science, Vol.5, No.2, pp 623- 627.

[4]. Sah, B.K. (2013): Generalisations of some countable and continues mixtures of Poisson distribution and their applications , Ph.D. Thesis, Patna University , 40-46