J. KSIAM Vol.12, No.2, 81-95, 2008

ON ESTIMATION OF NEGATIVE POLYA-EGGENBERGER DISTRIBUTION AND ITS APPLICATIONS

ANWAR HASSAN[1†] AND SHEIKH BILAL[2]

[1]DEPARTMENT OF STATISTICS, UNIVERSITY OF KASHMIR, SRINAGAR INDIA
E-Mail address: anwar.hassan2007@gmail.com
[2]DEPARTMENT OF STATISTICS, DEGREE COLLEGE, BARAMULLA KASHMIR INDIA
E-Mail address: sbilal_sbilal@yahoo.com

ABSTRACT. In this paper, the negative Polya-Eggenberger distribution has been introduced by compounding negative binomial distribution with beta distribution of I-kind which generates a number of univariate contagious or compound (or mixture of) distributions as its particular cases. The distribution is unimode, over dispersed and all of its positive and negative integer moments exist. The difference equation of the proposed model shows that it is a member of the Ord's family of distribution. Some of its interesting properties have been explored besides different methods of estimation been discussed. Finally, the parameters of the proposed model have been estimated by using a computer programme in R-software. Application of the proposed model to some data, available in the literature, has been given and its goodness of fit demonstrated.

1. INTRODUCTION

The Polya-Eggenberger distribution and its inverse negative Polya-Eggenberger distribution was introduced by Polya and Eggenberger [1] through an urn model. They described these distributions as truly contagious distribution. The other type such as Neyman's contagious distribution [2] involves the "apparent contagion" as described by Feller [3]. The distribution in discussion has not been studied in detail so for. Though, its different parameterization is present in the literature. The first parameterization of the distribution is the beta-negative binomial distribution which is the generalization of hyper-geometric distribution; see Johnson and Kotz's [10] for details. Another parameterization is the generalized Waring distribution introduced by Irwin [9] as an accident proneness-liability model which was subsequently applied by Irwin {[9], [14]} to data on accidents sustained by men in a soap factory. The negative Polya-Eggenberger distribution also belongs to Kemp and Kemp [6] generalized hyper-geometric distribution.

The negative Polya-Eggenberger distribution is related to Polya-Eggenberger

Received by the editors January 18, 2008.
2000 *Mathematics Subject Classification. 62E15.*
Key words and phrases. Negative binomial distribution, negative Polya-Eggenberger distribution, recurrence relations, probability generating function, chi-square fitting.
† Corresponding author.

distribution in the same way as negative binomial distribution is related to binomial distribution. It is well known fact that the negative binomial distribution 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 stasfied.There are various extensions/modifications of NBD in the literature including Engen's extended NBD {[13], [15]}, GNBD of Jain and Consul [12] and Weighted NBD; see Johnson et al. [16] for more details and explanations.

In this paper, the proposed model has been obtained by compounding the negative binomial distribution with beta distribution of first kind. In Gurland's [7] terminology, the proposed model represents a generalization of the negative binomial distribution. It has been shown that the proposed model exhibits more flexible alternative model to negative-binomial distribution and some of its generalizations. This has been demonstrated with the help of three data sets by the goodness of fit in the last section of this paper.

The organization of the paper is as follows: in section 2, we derive the proposed model. Section 3 deals with some interesting structural properties of the proposed model. In section 4, we discuss its relation with other distributions. Section 5 proposes different methods of estimation. Finally, in section 6, the parameters of the proposed model have been estimated by using a computer programme in R-soft wear. Further, an application of the proposed model to three data sets has been given and its goodness of fit demonstrated.

2. THE PROPOSED MODEL

A certain mixture distribution arises when all (or some) parameters of a distribution vary according to some probability distribution called the mixing distribution. A well-known example of discrete-type mixture distribution is the negative-binomial distribution which can be obtained as a Poisson mixture with gamma distribution.

Let X has a conditional negative-binomial distribution with parameter p, that is, X has a conditional probability mass function (pmf)

$$P(X/p)=P(X=x/p)=\binom{n+x-1}{x}p^x(1-p)^n \quad \text{for } x=0,1,2,.. \text{ and } 0<p<1,\ n>0 \quad (2.1)$$

Now, suppose p is a continuous random variable with probability density function (pdf)

$$g(p)=\frac{1}{\beta(\gamma,\alpha)}p^{\gamma-1}(1-p)^{\alpha-1} \quad \text{for} \quad 0<p<1, \quad (\alpha,\gamma)>0 \qquad (2.2)$$

Bhattacharya [8] showed that the conditional pmf of X is given by

$$f(x)=P(X=x)=\int_0^\infty f(x/p)\,g(p)dp$$

The equation above together with (2.1) and (2.2) gives

$$P(X=x)=\binom{n+x-1}{x}\frac{\alpha(\alpha+1).......(\alpha+\overline{x-1})\,\gamma(\gamma+1).........(\gamma+\overline{n-1})}{(\alpha+\gamma)(\alpha+\gamma+1)............(\alpha+\gamma+\overline{n+x-1})}$$

Taking $\alpha=a/c$, $\gamma=b/c$, the equation above reduces to the negative Polya-Eggenberger distribution with pmf

$$P(X=x)=\binom{n+x-1}{x}\frac{a(a+c).......(a+\overline{x-1}c)b(b+c).........(b+\overline{n-1}c)}{(a+b)(a+b+c)............(a+b+\overline{n+x-1}c)} \quad \text{for} \quad x=0,1,2,.. \quad (2.3)$$

The proposed model (2.3) can be put into different forms for the mathematical convenience and to study some of its properties. The model (2.3) in terms of ascending factorials can be put as

$$P(X=x)=\binom{n+x-1}{x}\frac{a^{[x,c]}\ b^{[n,c]}}{(a+b)^{[n+x,c]}}\ , \qquad \text{for} \quad x=0,1,2,....... \qquad (2.4)$$

Where $a^{[x,c]}=a(a+c)..........(a+\overline{x-1}\ c)$

Another form of (2.3) can be

$$P(X=x)=\binom{n+x-1}{x}\frac{\alpha^{[x]}\ \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}}\ , \qquad \text{for} \quad x=0,1,2,........... \qquad (2.5)$$

Where $\alpha^{[x]}=\alpha(\alpha+1)........(\alpha+\overline{x-1})$ and $\alpha=a/c$, $\gamma=b/c$. The model represented by (2.5) has been seen the most workable model, used through out this paper, for the mathematical computations.

Another form of (2.3) in terms of $n,\ p=\dfrac{a}{(a+b)}$, $Q=(1-p)=\dfrac{b}{(a+b)}$ and $\delta=\dfrac{c}{(a+b)}$ can be

$$P(X=x)=\binom{n+x-1}{x}\frac{\prod_{j=0}^{x-1}(p+j\delta)\prod_{j=0}^{n-1}(Q+j\delta)}{\prod_{j=0}^{n+x-1}(1+j\delta)}\ , \quad \text{for} \quad x=0,1,2,........... \qquad (2.6)$$

REMARKS: A number of special cases can be deduced from the proposed model (2.3) by assigning different set of values to its parameters. Some of the interesting cases deduced are negative-binomial distribution, beta–negative binomial distribution, negative hyper geometric distribution, geometric series distribution, Bernoulli-delta distribution (geometric) etc.

3. STRUCTURAL PROPERTIES

In this section, some of the interesting properties of the proposed model has been explored which are described as follows;

3.1 RECURRENCE RELATION BETWEEN PROBABILITIES

Expressing the pmf of the proposed model (2.5) as

$$P(X=x)=\frac{(n+x-1)!}{(n-1)!\ x!}\ \frac{\alpha^{[x]}\quad \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}} \tag{3.1}$$

Taking $x=x+1$ in the equation above and dividing the resulting equation by (3.1), we get the recurrence relation of the proposed model as

$$P(X=x+1)=\left[\frac{(n+x)}{(x+1)}\ \frac{(\alpha+x)}{(\alpha+\gamma+n+x)}\right]P(X=x) \tag{3.2}$$

Which yields the difference equation of the proposed model as

$$\Delta P_{x-1}=\frac{\left\{\dfrac{n\alpha-n-\alpha}{\gamma+1}-x\right\}P_x}{\left\{\dfrac{n\alpha-n-\alpha}{\gamma+1}+\dfrac{1}{\gamma+1}\right\}+\left\{\dfrac{n+\alpha+\gamma}{\gamma+1}-1\right\}x+\dfrac{x(x+1)}{\gamma+1}}$$

The difference equation above exhibits that the proposed model is a member of the Ord's family of distribution.

3.2. UNIMODALITY

The proposed model is a unimodal by the following result of Holgate [11]:

LEMMA. *If the mixing distribution is non-negative, continuous and unimodal then the resulting distribution is unimodal*

The proposed model is a unimodal since the mixing distribution is a beta distribution of I-kind which is unimodal. To show the unimodality of the distribution we have the following theorem.

Theorem 3.1. The proposed model is a unimodal for all values of (n,α,γ) and the mode is at $x=0$ if $n\alpha<1$ and for $n\alpha>1$ the mode is at some other point $x=M$ such that

$$\frac{n(\alpha-1)-(\alpha+\gamma)}{(\gamma+1)}<M<\frac{(n-1)(\alpha+1)}{(\gamma+1)} \tag{3.3}$$

Proof. The recurrence relation (3.2) gives the ratio

$$\frac{P(x+1)}{P(x)}=\frac{(n+x)}{(x+1)}\ \frac{(\alpha+x)}{(\alpha+\gamma+n+x)} \tag{3.4}$$

Which is less than one, that is,

$$\frac{P(x+1)}{P(x)}<1 \text{ if } n\alpha<1 \quad \forall (n,\alpha,\gamma)>0$$

Hence, for $n\alpha<1$, the ratio $\frac{P(x+1)}{P(x)}$ is a non-increasing function, therefore, the mode of the proposed model exists at $x=0$. Suppose for $n\alpha>1$ the mode exists at $x=M$, then the ratio defined by (3.4) gives the two inequalities

$$\frac{P(M+1)}{P(M)}=\frac{(n+M)}{(M+1)}\ \frac{(\alpha+M)}{(\alpha+\gamma+n+M)}<1 \tag{3.5}$$

and

$$\frac{P(M)}{P(M-1)}=\frac{n+M-1}{M}\frac{(\alpha+M-1)}{(\alpha+\gamma+n+M-1)}>1 \tag{3.6}$$

By the inequality (3.5) we have

$$\frac{n(\alpha-1)-(\alpha+\gamma)}{(\gamma+1)}<M \tag{3.7}$$

and the inequality (3.6) gives

$$M<\frac{(n-1)(\alpha+1)}{(\gamma+1)} \tag{3.8}$$

On combining the inequalities (3.7) and (3.8), we get the result (3.3).

3.3. MEAN AND VARIANCE

Mean and variance of the proposed model can be easily obtained by using the properties of conditional mean and variance as follows;

MEAN: By the conditional mean we have

$$Mean\ =E(X)=E[E(X/p)] \tag{3.9}$$

Where $E(X/p)$ is the conditional expectation of X given p and for given p the random variable X follows (2.1) with mean and variance given by

$$\left.\begin{aligned} E(X/p)&=np^{-1}(1-p) \\ V(X/p)&=np^{-2}(1-p) \end{aligned}\right\} \tag{3.10}$$

The equation (3.9) together with (3.10) and (2.2) gives mean of the proposed model as

$$E(X)=\frac{n\alpha}{(\gamma-1)} \qquad \gamma>1$$

VARIANCE: Similarly, by the conditional variance we have

$$V(X)=E[V(X/p)]+V[E(X/p)] \tag{3.11}$$

Using (3.10) in the equation above, we get

$$V(X)=nE[p^{-2}(1-p)]+n^2E\left[p^{-2}(1-p)^2\right]-n^2\{E(p^{-1}(1-p))\}^2$$

Since p is varying as (2.2), the equation above reduces to

$$V(X)=\frac{n}{\beta(\gamma,\alpha)}\int_0^1 p^{\gamma+1-1}(1-p)^{\alpha-3}dp+\frac{n^2}{\beta(\alpha,\gamma)}\int_0^1 p^{\gamma+2-1}(1-p)^{\alpha-3}dp-\{E(X)\}^2$$

By an application of beta integral, the equation above gives variance as

$$V(X)=\frac{n\alpha}{(\gamma-1)}+\frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}-\left(\frac{n\alpha}{(\gamma-1)}\right)^2 \qquad \text{for} \quad \gamma>2$$

3.4. RECURRENCE RELATION BETWEEN MOMENTS

The recurrence relation (3.2) gives

$$(1+x)^{(r+1)}P_{x+1}(n,\alpha,\gamma)=(1+x)^{r}\frac{(\alpha+x)(n+x)}{(\alpha+\gamma+n+x)}P_{x}(n,\alpha,\gamma)$$

Which subsequently reduces to

$$(1+x)^{(r+1)}P_{x+1}(n,\alpha,\gamma)=(1+x)^{r}\frac{n\alpha}{(\gamma-1)}P_{x}(n+1,\alpha+1,\gamma-1),\quad \gamma>1 \tag{3.12}$$

Where $P_{x}(n+1,\alpha+1,\gamma-1)$ denotes the pmf of the proposed model with parameters$(n+1,\alpha+1,\gamma-1)$. Summing (3.12) over the values of x on both sides, we get the moment recurrence relation as

$$\mu'_{r+1}(n,\alpha,\gamma)=\frac{n\alpha}{(\gamma-1)}\sum_{j=0}^{r}\binom{r}{j}\mu'_{j}(n+1,\alpha+1,\gamma-1) \tag{3.13}$$

Where $\mu'_{j}(n+1,\alpha+1,\gamma-1)$ denotes the jth moment about origin of the proposed model with parameters$(n+1,\alpha+1,\gamma-1)$. The recurrence relation (3.13) gives the first four moments about origin as

$$\mu'_1=\frac{n\alpha}{(\gamma-1)},\qquad \gamma>1$$

$$\mu'_2=\frac{n\alpha}{(\gamma-1)}+\frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)},\qquad \gamma>2$$

$$\mu'_3=\frac{n\alpha}{(\gamma-1)}+\frac{3n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}+\frac{n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)},\qquad \gamma>3$$

$$\mu'_4=\frac{n\alpha}{(\gamma-1)}+\frac{7n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}+\frac{6n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)}$$
$$+\frac{n(n+1)(n+2)(n+3)\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma-1)(\gamma-2)(\gamma-3)(\gamma-4)},\qquad \gamma>4$$

Now, the central moments can be easily obtained from the moments about origin of the proposed model and are given by

$$\mu_2=\frac{n\alpha}{(\gamma-1)}\left[1+\frac{(n+1)\,(\alpha+1)}{(\gamma-2)}-\frac{n\alpha}{(\gamma-1)}\right],\qquad \gamma>2$$

$$\mu_3=\frac{n\alpha}{(\gamma-1)}\left[1-\frac{3n\alpha}{(\gamma-1)}+2\left(\frac{n\alpha}{(\gamma-1)}\right)^{2}\right]+\frac{3n(n+1)\,\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}\left[1-\frac{n\alpha}{(\gamma-1)}\right]$$
$$+\frac{n(n+1)(n+2)\ \alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)},\qquad \gamma>3$$

$$\mu_4=\frac{n\alpha}{(\gamma-1)}\left[1-\frac{4n\alpha}{(\gamma-1)}+6\left(\frac{n\alpha}{(\gamma-1)}\right)^{2}-3\left(\frac{n\alpha}{(\gamma-1)}\right)^{3}\right]$$
$$+\frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}\left[7-\frac{12n\alpha}{(\gamma-1)}+6\left(\frac{n\alpha}{(\gamma-1)}\right)^{2}\right]$$

$$+\frac{n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)}\left[6-\frac{4n\alpha}{(\gamma-1)}\right]$$
$$+\frac{n(n+1)(n+2)(n+3)\alpha(\alpha+1)(\alpha+2)(\alpha+3)}{(\gamma-1)(\gamma-2)(\gamma-3)(\gamma-4)} \quad , \gamma>4$$

The proposed model is over dispersed for $\gamma>2$ and its coefficient of variation is given by

$$C.V.=1+\frac{(n+1)(\alpha+1)}{(\gamma-2)}-\frac{n\alpha}{(\gamma-1)}, \qquad \gamma>2$$

3.5. PROBABILITY GENERATING FUNCTION

Suppose $G_X(u)$ denotes the probability generating function of the proposed model then we have

$$G_X(u) = E(u^x)=\sum_{x=0}^{\infty} u^x \frac{(n+x-1)!}{(n-1)!\, x!} \frac{\alpha^{[x]}\ \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}}$$

Which yields the probability generating function of the proposed model as

$$G_X(u) = \frac{\gamma^{[n]}}{(\alpha+\gamma)^{[n]}}\ {}_2F_1[n,\alpha\ ;\alpha+\gamma+n,u]$$

Where ${}_2F_1[n,\alpha\ ;\alpha+\gamma+n,u]$ is a Gaussian hypergeometric function.

REMARKS: If we replace u by $(1+t)$ and $(1-t)$, the equation above yields the descending factorial moment and ascending factorial moment generating functions respectively.

3.6. FACTORIAL MOMENTS

The rth factorial moment about origin $\mu'_{(r)}$ of the proposed model is defined as

$$\mu'_{(r)} = E(X^{(r)})=\sum_{x=0}^{\infty} x^{(r)} \frac{(n+x-1)!}{(n-1)!\ x!} \frac{\alpha^{[x]}\ \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}}$$

Which reduces to

$$\mu'_{(r)}=\sum_{x=r}^{\infty} \frac{(n+x-1)!}{(n-1)!\,(x-r)!} \frac{\alpha^{[x]}\ \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}}$$

Taking $x=x+1$, the equation above yields the rth factorial moment of the proposed model as

$$\mu'_{(r)} = \frac{n^{[r]}\ \alpha^{[r]}\,(\gamma-r-1)!}{(\gamma-1)!} \tag{3.14}$$

The expression (3.14) can be used to obtain the factorial moments of the proposed model.

3.7. NEGATIVE INTEGER FACTORIAL MOMENTS (NIFM)

The negative moments are useful in many problems of applied statistics, especially in life testing and in survey sampling, where ratio estimates are used. In this section, we obtained the expression for the rth negative integer ascending factorial moment of the proposed model in terms of Gaussian hypergeometric function. Suppose $\varphi'_{-[r]}$ denotes the rth negative integer ascending factorial moment of the proposed mode then we have

$$\varphi'_{-[r]} = E[(x+1)^{[r]}]^{-1} = \sum_{x=0}^{\infty} \frac{1}{(x+1)^{[r]}} \frac{(n+x-1)!}{(n-1)!\ x!} \frac{\alpha^{[x]}\ \gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}} \tag{3.15}$$

Where $\frac{1}{(x+1)^{[r]}} == \frac{x!}{(x+r)!} = \frac{1^{[x]}}{1^{[r]}(r+1)^{[x]}}$

Using the result above in (3.15), we get

$$\varphi'_{-[r]} == \frac{\gamma^{[n]}}{r!(\alpha+\gamma)^{[n]}}\ {}_3F_2[n,\alpha,1;r+1,\alpha+\gamma+n;1]$$

Which gives the first four negative integer ascending factorial moment as

$$\varphi'_{-[1]} = \frac{\gamma^{[n]}}{(\alpha+\gamma)^{[n]}}\ {}_3F_2[n,\alpha,1;2,\alpha+\gamma+n;1]$$

$$\varphi'_{-[2]} = \frac{\gamma^{[n]}}{2(\alpha+\gamma)^{[n]}}\ {}_3F_2[n,\alpha,1;3,\alpha+\gamma+n;1]$$

$$\varphi'_{-[3]} = \frac{\gamma^{[n]}}{6(\alpha+\gamma)^{[n]}}\ {}_3F_2[n,\alpha,1;4,\alpha+\gamma+n;1]$$

$$\varphi'_{-[4]} = \frac{\gamma^{[n]}}{24(\alpha+\gamma)^{[n]}}\ {}_3F_2[n,\alpha,1;5,\alpha+\gamma+n;1]$$

4. RELATION WITH OTHER DISTRIBUTIONS

Theorem 4.1. *Let* X *be a negative Polya-Eggenberger variate with parameters* (n,α,γ) *.If* $\gamma\to\infty$ *such that* $\alpha\gamma^{-1}=\lambda$ *and* $\lambda=\theta n^{-1}$ *as* $n\to\infty$ *then show that* X *tends to a Poisson distribution with parameter* θ *.*

Proof: Expressing the pmf of the proposed model as

$$P(X=x) = \frac{(n+x-1)(n+x-2).......(n+1)n}{x!} \frac{\alpha(\alpha+1)...(\alpha+x-1)\gamma(\gamma+1)...(\gamma+n-1)}{(\alpha+\gamma)(\alpha+\gamma+1)...(\alpha+\gamma+n+x-1)} \tag{4.1}$$

Taking limit $\gamma\to\infty$ such that $\alpha\gamma^{-1}=\lambda$, the equation above gives

$$\underset{\gamma\to\infty}{lt}\ P(X=x) = \left(1+\frac{x-1}{n}\right)\left(1+\frac{x-2}{n}\right).....\left(1+\frac{1}{n}\right) \frac{(n\lambda)^x}{x!(1+\lambda)^{n+x}}$$

Substituting $\lambda=\frac{\theta}{n}$ and taking limit $n\to\infty$, the equation above reduces to the Poisson distribution with parameter θ.

Theorem 4.2. *Let* X *be a negative Polya-Eggenberger variate with parameters* (n,α,γ). *Show that zero-truncated negative Polya-Eggenberger distribution tends to logarithmic series distribution.*

Proof: The pmf of the Zero-truncated negative Polya-Eggenberger distribution is

$$P(X=x)=\binom{n+x-1}{x}\frac{\alpha^{[x]}\gamma^{[n]}}{(\alpha+\gamma)^{[n+x]}}\frac{(\alpha+\gamma)^{[n]}}{(\alpha+\gamma)^{[n]}-\gamma^{[n]}}\quad,\quad x=1,2,........ \tag{4.2}$$

Substituting $\alpha\gamma^{-1}=\lambda$ and proceeding to limit $\gamma\to\infty$, we get

$$\underset{\gamma\to\infty}{lt}P(X=x)=\frac{n\Gamma(n+x)}{\Gamma(n+1)\Gamma(x+1)}\frac{\lambda^x}{(1+\lambda)^{n+x}}\frac{1}{1-(1+\lambda)^{-n}}$$

Taking $\frac{\lambda}{1+\lambda}=t$ in the equation above, we get

$$P(X=x)=\frac{n\Gamma(n+x)}{\Gamma(n+1)\Gamma(x+1)}\frac{t^x(1-t)^n}{1-(1-t)^n}$$

Proceeding to the limit $n\to 0$, the equation above reduces to the logarithmic series distribution.

5. ESTIMATION

5.1. MOMENT METHOD:

Let m_1',m_2',m_3' be the sample moments about origin and μ_1',μ_2',μ_3' be the population moments about origin of the proposed model. The method of moments consists in comparing the sample moments with the population moments of the proposed model, that is,

$$\bar{x}=\frac{n\alpha}{(\gamma-1)} \tag{5.1}$$

$$s^2+\bar{x}^2=\frac{n\alpha}{(\gamma-1)}+\frac{n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)},\qquad sample\ variance=s^2=m_2'-\bar{x}^2 \tag{5.2}$$

Using (5.1) in (5.2), we get

$$\frac{l_1(\gamma-2)-\bar{x}^2(\gamma-1)}{\bar{x}}=n+\alpha+1\ ,\qquad l_1=s^2+\bar{x}^2-\bar{x} \tag{5.3}$$

By comparing the third sample moment with its corresponding population moment, we get

$$m_3'=\frac{n\alpha}{(\gamma-1)}+\frac{3n(n+1)\alpha(\alpha+1)}{(\gamma-1)(\gamma-2)}+\frac{n(n+1)(n+2)\alpha(\alpha+1)(\alpha+2)}{(\gamma-1)(\gamma-2)(\gamma-3)} \tag{5.4}$$

The equation (5.4) together with (5.1) and (5.2) gives

$$\frac{(\gamma-3)(m_3'-\bar{x}-3l_1)}{l_1}-\frac{(\gamma-2)l_1}{\bar{x}}=n+\alpha+3 \tag{5.5}$$

Eliminating n and α between (5.3) and (5.5), we get the estimate of γ as

$$\hat{\gamma}=\frac{\bar{x}(3m_3'-7l_1)-4l_1^2+\bar{x}^2(l_1-3)}{\bar{x}(m_3'-3l_1)-2l_1^2+\bar{x}^2(l_1-1)}$$

Substituting the value of n from (5.3) into (5.1), we get a quadratic equation in α as

$$\alpha^2\bar{x}-\alpha[l_1(\hat{\gamma}-2)-\bar{x}^2(\hat{\gamma}-1)-\bar{x}]+\bar{x}^2(\hat{\gamma}-1)=0$$

Which can be solved for α. After estimating α, the value of n can be obtained either from (5.5) or (5.1).

5.2. USING MEAN AND FIRST THREE CELL FREQUENCIES:

Equating the first three probabilities of the proposed model with their corresponding relative frequencies $\frac{f_0}{N},\frac{f_1}{N},\frac{f_2}{N}$, we get

$$\frac{\gamma^{[n]}}{(\alpha+\gamma)^{[n]}}=\frac{f_0}{N} \tag{5.6}$$

$$\frac{n\alpha\gamma^{[n]}}{(\alpha+\gamma)^{[n+1]}}=\frac{f_1}{N} \tag{5.7}$$

$$\frac{n(n+1)\alpha(\alpha+1)\gamma^{[n]}}{2(\alpha+\gamma)^{[n+2]}}=\frac{f_2}{N} \tag{5.8}$$

Dividing (5.6) by (5.7) and then using (5.1) in the resulting equation, we get

$$\gamma=\frac{tf_1+\bar{x}f_0}{\bar{x}f_0} \tag{5.9}$$

Where $\alpha+\gamma+n=t$ (5.10)

Using (5.1) and (5.10) in the equation obtained by dividing (5.8) with (5.7), we get

$$\gamma=\frac{2(t+1)f_2-(t+1-\bar{x})f_1}{(\bar{x}-1)f_1} \tag{5.11}$$

Eliminating γ between (5.9) and (5.11), we get the estimate of t as

$$\hat{t}=\frac{2\bar{x}f_0f_2}{f_1^2(\bar{x}-1)-2\bar{x}f_0f_2+\bar{x}f_0f_1}$$

The equation (5.9) together with the result obtained above gives the estimate of γ as

$$\hat{\gamma}=\frac{\hat{t}f_1+2\bar{x}f_0}{\bar{x}f_0}$$

Substituting the value of n from (5.10) into (5.1), we get a quadratic equation in α as

$$\alpha^2-\alpha(\hat{t}-\hat{\gamma})+\bar{x}(\hat{\gamma}-1)=0$$

Which can be used to estimate α. The estimate of n can be obtained from (5.10) or (5.1).

5.3. MAXIMUM LIKELIHOOD METHOD:

The log likelihood function of the proposed model is given by

$$log\,L = N\left[log(\gamma)^{[n]} - log(\alpha+\gamma)^{[n]}\right] + \sum f_x log(n)^{[x]}$$
$$+\sum f_x log(\alpha)^{[x]} - \sum f_x log(\alpha+\gamma+n)^{[x]} - \sum f_x log(x!)$$

Where f_x is the observed frequency for the variate value x and $N=\sum f_x$.

The method of maximum likelihood method of estimation gives the three likelihood equations for three unknown parameters as

1) $\dfrac{\partial log\,L}{\partial\alpha} = 0 = \sum f_x \sum_{k=0}^{n+x-1} \dfrac{1}{\alpha+\gamma+k} - \sum f_x \sum_{k=0}^{x-1} \dfrac{1}{\alpha+k}$

2) $\dfrac{\partial log\,L}{\partial\gamma} = 0 = N \sum_{k=0}^{n-1} \dfrac{1}{\gamma+k} - \sum f_x \sum_{k=0}^{n+x-1} \dfrac{1}{\alpha+\gamma+k}$

3) $\dfrac{\partial \log L}{\partial n} = 0 = N\dfrac{\partial \log(\gamma)^{[n]}}{\partial n} - N\dfrac{\partial \log(\alpha+\gamma)^{[n]}}{\partial n}$

$$+\sum f_x \frac{\partial \log(n)^{[x]}}{\partial n} - \sum f_x \frac{\partial \log(\alpha+\gamma+n)^{[x]}}{\partial n}$$

Converting the likelihood equation above in terms of gamma functions, we get

$$N\frac{\partial log\,\Gamma(\gamma+n)}{\partial n} - N\frac{\partial log\,\Gamma(\alpha+\gamma+n)}{\partial n} + \sum f_x \frac{\partial log\,\Gamma(n+x)}{\partial n} - \sum f_x \frac{\partial log\,\Gamma(n)}{\partial n}$$
$$+\sum f_x \frac{\partial log\,\Gamma(\alpha+\gamma+n+x)}{\partial n} - \sum f_x \frac{\partial log\,\Gamma(\alpha+\gamma+n)}{\partial n} = 0 \qquad (5.12)$$

The differentiation of the equation above is not straight forward and can be solved with the help of the following recurrence relation; (see pages 6-8 Johnson, Kotz and Kemp [16] for details)

$$\psi(x+n) = \psi(x) + \sum_{j=1}^{n} (x+j-1)^{-1}, \qquad n=1,2,..... \qquad (5.13)$$

Where $\psi(x) = \dfrac{d}{dx}\{log\,\Gamma(x)\} = \dfrac{\Gamma'(x)}{\Gamma(x)}$ is called digamma function. A good approximation for $\psi(x)$ is

$$\psi(x) = log\,\Gamma(x) - \frac{1}{2x}$$

The relation (5.13) together with the above result gives

$$\psi(x+n) = log\,\Gamma(x) - \frac{1}{2x} + \sum_{j=1}^{n} (x+j-1)^{-1}, \qquad n=1,2,..... \qquad (5.14)$$

By an application of (5.14), the equation (5.12) gives the third kelihood equation as

$$N\left[log\,\Gamma(\gamma+n) + \sum_{k=1}^{\gamma} (n+k-1)^{-1}\right] - 2N\left[log\,\Gamma(\alpha+\gamma+n) + \sum_{k=1}^{\alpha+\gamma} (n+k-1)^{-1}\right]$$
$$+\sum f_x\left[log\,\Gamma(n+x) + \sum_{k=1}^{x} (n+k-1)^{-1}\right] - Nog\Gamma(n)$$

$$+\sum f_x\left[\log\Gamma(n+\alpha+\gamma+x)+\sum_{k=1}^{\alpha+\gamma+x}(n+k-1)^{-1}\right]=0$$

The three likelihood equations are not simple to provide direct solution, however, different iterative procedures such as Fisher's scoring method, Newton-Rampson method etc. can be employed to solve these equations. We may solve the following system of equations

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

Where $\hat{\theta}=(n,\alpha,\gamma)$ is a parameter vector, is the ML estimate of θ and θ_0 is the trial value of θ which may be first obtained by equating the theoretical frequencies with the observed frequencies.

6. CHI-SQUARE FITTING

In this section, we present three data sets to examine the fitting of the proposed model and comparing that with the negative binomial distribution and generalized negative binomial distribution defined by Jain and Consul [12].

As mentioned in the previous section, the maximum likelihood equations are not straightforward to provide the maximum likelihood (ML) estimates of the parameters of the proposed model and thus need some iterative procedure such as Fisher's scoring method, Newton-Rampson method etc. for their solution. The R-soft wear provides one among such solutions. The parameters have been estimated with the help of a computer program in R-soft wear and has been shown in the bottom of the table.

TABLE.1. Absenteeism among shift-workers in steel industry; data of Arbous and Sichel [5]

Count	Observed frequency	Expected frequencies		
		NBD	Jain and Consul's[12] GNBD	Proposed model NPED
0	7	12.02	10.51	9.53
1	16	16.16	17.45	15.93
2	23	17.77	20.38	19.06
3	20	18.08	20.80	19.92
4	23	17.65	19.88	19.41
5	24	16.80	18.34	18.17
6	12	15.72	16.56	16.59
7	13	14.52	14.78	14.90
8	09	13.28	13.08	13.25
9	09	12.06	11.53	11.71
10	08	10.89	10.13	10.30
11	10	09.78	08.89	9.04
12	08	08.75	07.79	7.92
13	07	07.80	16.83	6.94

14	02	06.93	05.99	6.08
15	12	06.14	05.26	5.33
16	03	05.43	04.61	4.68
17	05	04.79	04.05	4.12
18	04	04.22	03.56	3.63
19	02	03.17	03.14	3.20
20	02	03.23	02.76	2.83
21	05	02.86	02.43	2.50
22	05	02.50	02.15	2.22
23	02	02.91	01.90	1.97
24	01	01.91	01.68	1.75
25-48	16	12.77	13.50	17.02
TOTAL	248	248	248	248
ML Estimate		p=0.854 $n=1.576$	p=0.0001077 β =5978.528 n=29337.083	n=14.962954 α=2.492821 γ=4.852530
χ^2		14.92	27.79	10.20
d. f		17	16	16

TABLE 2. The data has been taken from Beall-Rescia [4], Table VII

Count	Observed frequency	Expected frequencies		
		NBD	Jain and Consul's[12] GNBD	Proposed model NPED
0	206	198.85	197.64	196.16
1	143	166.90	168.45	170.06
2	128	124.33	125.12	126.35
3	107	88.70	88.73	89.08
4	71	61.89	61.56	61.32
5	36	42.60	42.19	41.71
6	32	29.05	28.71	28.21
7	17	19.68	19.45	19.04
8	4	13.27	13.13	12.84
9	7	8.91	8.85	8.67
10	7	5.96	5.96	5.87
11	2	3.98	4.00	3.98
12	3	2.65	2.69	2.71
13	3	1.76	1.81	1.85
14	1	1.17	1.21	1.27
15	1	0.78	0.82	0.87
16	1	0.51	0.55	0.60
17	2	0.34	0.37	0.42
18	1	0.22	0.25	0.29

19	0	0.45	0.51	0.70
TOTAL	772	772	772	772
ML Estimate		n=1.2903314 p=0.6504948	n=1.4392633 β =1.0347577 p=0.6119716	n= 45.146373 α=1.429784 γ=27.880158
χ^2		18.27	18.18	17.68
d. f		9	8	8

TABLE 3. Accidents to 647 women working on H.E. Shells during 5 weeks

Count	Observed frequency	Expected frequencies		
		NBD	Jain and Consul's[12] GNBD	Proposed model NPED
0	447	445.89	445.17	445.58
1	132	134.90	136.76	135.88
2	42	44.59	43.09	43.41
3	21	14.09	14.27	14.40
4	3	4.96	4.93	4.94
5	2	2.57	2.78	2.79
TOTAL		647	647	647
ML Estimate		n= 0.86512 p=0.34969	n= 8.26367 β =4.84301 p=0.04424	n=17.78681 α= 0.98208 γ=38.51329
χ^2		4.46	4.32	4.15
d. f		2	1	1

From all the tables it is clear that the proposed model gives a very close fit as compared to other distributions. Thus, the proposed model provides a better alternative to explain the data than the compared distributions.

ACKNOWLEDGMENTS

The authors are thankful to the referees for useful suggestions which improved the quality and presentation of this paper.

REFERENCES

[1] F. Eggenberger and G. Polya, *ZAMM*: *Z. Angew. Math. Mech,* Uber die Statistik verketteter vorgange, 1(1923), 279-289.

[2] J. Neyman, *AMS: Ann. Math. Statist*., On a new class of contagious distribution applicable in entomology and bacteriology, 10 (1939), 35-57.

[3] W. Feller, *AMS: Ann. Math. Statist.,* On a general class of contagious distributions, 14 (1943), 389-400.

[4] G. Beall and R. R. Rescia, *B: Biometrics*, A generalization of Neyman's contagious distributions, 9(1953), 354-57.

[5] A. G. Arbous and H. S. Sichel, *B: Biometrika,* New techniques for the analysis of absenteeism data, 41(1954), 77.

[6] C. D. Kemp and A. W. Kemp., JRSS: *Journal of the Royal Statistical Society*, *Series B,* Generalized hypergeometric distributions, 18, (1956 a), 202-211.

[7] J. Gurland, *B: Biometrics,* A generalized class of contagious distribution, 14 (1957), 229-249.

[8] S. K. Bhattacharya, *CSAB: Calcutta Statist. Assoc. Bull.,* Confluent hypergeometric distributions of discrete and continuous type with applications to accident proneness, 15(1966), 20-31.

[9] J. O. Irwin, JRSS: *Journal of the Royal Statistical Society*, *Series A,* The generalized Waring distribution applied to accident theory, 131(1968), 205-225.

[10] N. L. Johnson and S. Kotz, *Univariate Discrete Distribution,* New York: John Wiley and Sons, (1969).

[11] P. Holgate, *B: Biometrika,* The modality of compound Poisson distribution, 57(1970), 665-667.

[12] G. C. Jain and P. C. Consul, *SJAM: SIAM J. Appl. Math*, A Generalized Negative Binomial Distribution, 21(4) (1971), 501-513.

[13] S. Engen, *B: Biometrika,* On species frequency models, 61(1974), 263-270.

[14] J. O. Irwin, JRSS: *Journal of the Royal Statistical Society*, *Series A,* The generalized Waring distribution Part II, 138(1975 b), 204-227.

[15] S. Engen, *Stochastic Abundance Model.* London: Chapman and Hall, (1978).

[16] N.L. Johnson, S. Kotz, and A. W. Kemp, *Univariate Discrete distributions,* New York: John Wiley and Sons, (1993).