J. KSIAM Vol.11, No.2, 15-29, 2007

# ON SOME MODELS LEADING TO QUASI-NEGATIVE-BINOMIAL DISTRIBUTION

By

**SHEIKH BILAL**
Department of Statistics
Degree College, Baramula,
Kashmir, India
**Email-** sbilal_sbilal@yahoo.com

**And**

**ANWAR HASSAN**
PG Department of Statistics
University of Kashmir
Srinagar-India
**Email-** anwar_husan@yahoo.com

Abstract In this paper, we explore some interesting models of the quasi-negative-binomial distribution based on difference differential equations applicable to theory of microorganisms and the situations like that. Some characterizations based on conditional distributions and damage process have been obtained. Further, the distribution of number of accidents as the quasi-negative-binomial distribution in the light of Irwin's theory of "proneness-liability" model has been derived. Finally, the proposed model (QNBD) has been applied to study the Shunting accidents, home injuries, and strikes in industries.

Keywords: quasi-negative-binomial distribution, models based on difference differential equations, characterization, distribution of number of accidents, chi-square fitting.

## 1. INTRODUCTION AND MOTIVATION:

Unlike Jain and Consul's (1973) GPD model, the quasi-negative-binomial distribution also reveals the fact that the probability of success from trial to trial does not remain constant. However, in the real world of living beings the value of the probability changes according to the circumstances. These changes may be due to the inheritance of genes, psychological effects, feelings of social togetherness, previous experience, determination for successor or to face a common danger, adjustments needed for changes in environments, wisdom etc. In such cases, the classical negative binomial distribution does not fit well the data arising from these cases. In fact, these are situations where the probability does remain linearly dependent on the number of successes.

The quasi-negative-binomial distribution is an interesting distribution and has not been studied in detail so for. It was obtained in different forms by Janardan (1975), Nandi and Das (1994) and Sen and Jain (1996).

In this paper, we applied the proposed model to study microorganisms and obtained some models based on difference differential equations leading to the quasi-negative-binomial distribution. This has been shown in Section 2. Section 3 deals with some characterizations based on conditional distribution and damage process. The distribution of number of accidents as quasi-negative-binomial distribution in the light of Irwin's theory of "proneness-liability" model has been derived in Section 4. Finally, in Section 5, we applied the proposed model to study the Shunting accidents, home injuries, and strikes in industries and obtained some remarkable fit than GPD model.

## 2. Models Based on Difference Differential Equations

In a paper by Tukey(1949), the probability distribution of balls in boxes is a Poisson distribution while assuming that the probability $II_x$ of finding $x$ balls in a box is a function of mean number $\lambda$ of balls in the boxes such that $II_0 = 1$ for $\lambda = 0$, $\frac{dII_0}{d\lambda} = -II_0$ and $\frac{dII_x}{d\lambda} = -II_{x-1}$ for $x \geq 1$.This result was generalized by Consul (1988) for generalized Poisson distribution and proved two results for the generalization of Tukey's result. Consul (1990d) also proved these results for quasi-binomial distribution.

It has been shown by many authors that the classical negative binomial distribution has become increasingly more successful and more flexible alternative than Poisson distribution in accounting the data especially arising in the study of entomology, bacteriology, ecology etc. Taking this fact into consideration and noting that quasi-negative-binomial distribution is a generalization of Jain and Consul's (1973) generalized Poisson distribution, in Gurland's (1957) terminology, and that of classical negative-binomial distribution, we had made an attempt here to prove these results for the proposed model.

Let there be an infinite but countable number of available spaces for insects, bacteria, viruses or microbes. Let $\theta_1$ be the probability of initial desire of each one of them to get into a particular location. This value of $\theta_1$ may increase or decrease by a small quantity $\theta_2$ due to some factors like psychological effects, feeling of social togetherness, mutual consultations, communications, determination, prevalent conditions, and the numbers succeeding to get in that location. Let $M_x(a,\theta_1,\theta_2)$ be the probability of finding exactly '$a$'number of microbes in that location which will be function of $a$, $\theta_1$ and $\theta_2$. By changing each one of the parameters $\theta_1$ and $\theta_2$ we get the following two theorems.

***Theorem. 2.1.*** If the mean $\mu(a,\theta_1,\theta_2)$ for the probability distribution of microorganisms is increased by changing the parameter $\theta_1$ to $\theta_1 + \Delta\theta_1$ in such a manner that

$$\frac{\partial}{\partial\theta_1}M_0(a,\theta_1,\theta_2)=-\frac{a}{(1+\theta_1)}M_0(a,\theta_1,\theta_2) \tag{2.1}$$

and

$$\frac{\partial}{\partial\theta_1}M_x(a,\theta_1,\theta_2)=-\frac{(a+x)}{(1+\theta_1+x\theta_2)}M_x(a,\theta_1,\theta_2)+aM_{x-1}(a+1,\theta_1+\theta_2,\theta_2) \tag{2.2}$$

For all integral values of $x>0$ with the initial condition $M_0(a,0,\theta_2)=1$ and $M_x(a,0,\theta_2)=0$ for $x>0$, then the probability model $M_x(a,\theta_1,\theta_2)$ is the QNBD model $P_x(a,\theta_1,\theta_2)$.

***Proof.*** Equation (2.1) is a linear differential equation with integrating factor $(1+\theta_1)^a$ and the general solution $M_x(a,\theta_1,\theta_2)=C_0(1+\theta_1)^{-a}$. By making use of the initial condition $M_0(a,0,\theta_2)=1$, the constant $C_0=1$ and therefore,

$$M_0(a,\theta_1,\theta_2)=(1+\theta_1)^{-a}=P_0(a,\theta_1,\theta_2)$$

Taking $x=1$ in (2.2) and using the result above, we get the linear differential equation

$$\frac{\partial}{\partial\theta_1}M_1(a,\theta_1,\theta_2)=-\frac{(a+1)}{(1+\theta_1+\theta_2)}M_1(a,\theta_1,\theta_2)+\frac{a}{(1+\theta_1+\theta_2)^{a+1}}$$

with the integrating factor $(1+\theta_1+\theta_2)^{a+1}$ and the general solution

$$(1+\theta_1+\theta_2)^{a+1}M_1(a,\theta_1,\theta_2)=a\theta_1+C_1$$

By making use of the initial condition $M_1(a,0,\theta_2)=0$, the constant $C_1=0$ and

$$M_1(a,\theta_1,\theta_2)=a\theta_1(1+\theta_1+\theta_2)^{-a-1}=P_1(a,\theta_1,\theta_2)$$

Using the result above in (2.2) for $x=2$, we get

$$\frac{\partial}{\partial\theta_1}M_2(a,\theta_1,\theta_2)=-\frac{(a+2)}{(1+\theta_1+2\theta_2)}M_2(a,\theta_1,\theta_2)+\frac{a(a+1)(\theta_1+\theta_2)}{(1+\theta_1+2\theta_2)^{a+2}}$$

On multiplying the differential equation above with integrating factor $(1+\theta_1+2\theta_2)^{a+2}$ and then integrating w.r. to $\theta_1$, we get

$$(1+\theta_1+2\theta_2)^{a+2}M_2(a,\theta_1,\theta_2)=a(a+1)\left[\frac{\theta_1^2}{2}+\theta_1\theta_2\right]+C_2$$

By making use of the initial condition for $x=2$, we get $C_2=0$ and

$$M_2(a,\theta_1,\theta_2)=\frac{a(a+1)}{2!}\frac{\theta_1(\theta_1+2\theta_2)}{(1+\theta_1+2\theta_2)^{a+2}}=P_2(a,\theta_1,\theta_2) \tag{2.3}$$

Now, for $x=3$, equation (2.2) together with relation (2.3) gives

$$\frac{\partial}{\partial\theta_1}M_3(a,\theta_1,\theta_2)=-\frac{(a+3)}{(1+\theta_1+3\theta_2)}M_3(a,\theta_1,\theta_2)+\frac{a(a+1)(a+2)}{2!}\frac{(\theta_1+\theta_2)(\theta_1+3\theta_2)}{(1+\theta_1+3\theta_2)^{a+3}}$$

The general solution of the linear differential equation above is

$$(1+\theta_1+3\theta_2)^{a+3}M_3(a,\theta_1,\theta_2)=\frac{a(a+1)(a+2)}{3!}(\theta_1+3\theta_2)^2+C_3$$

Again, by making use of the initial condition $M_3(a,0,\theta_2)=0$, the constant $C_3$ becomes zero and we get

$$M_3(a,\theta_1,\theta_2)=\frac{a(a+1)(a+2)}{3!}\frac{\theta_1(\theta_1+3\theta_2)^2}{(1+\theta_1+3\theta_2)^{a+3}}=P_3(a,\theta_1,\theta_2)$$

Thus we have proved that $M_x(a,\theta_1,\theta_2)=P_x(a,\theta_1,\theta_2)$ for $x=0,1,2,3$. Now, it can be easily shown by the method of induction that

$M_x(a,\theta_1,\theta_2)=\dfrac{(a+x-1)!}{(a-1)!\;x!}\dfrac{\theta_1(\theta_1+x\theta_2)^{x-1}}{(1+\theta_1+x\theta_2)^{a+x}}=P_x(a,\theta_1,\theta_2)$ for all nonnegative integral values of $x$.

***Theorem.* 2.2.** If the mean $\mu(a,\theta_1,\theta_2)$ for the probability distribution of microorganisms is increased by changing the parameter $\theta_2$ to $\theta_2+\Delta\theta_2$ in such a manner that

$$\frac{\partial}{\partial\theta_2}M_0(a,\theta_1,\theta_2)=0 \tag{2.4}$$

and

$$\frac{\partial}{\partial\theta_2}M_x(a,\theta_1,\theta_2)=\frac{-x(a+x)}{(1+\theta_1+x\theta_2)}M_x(a,\theta_1,\theta_2)+\frac{a(x-1)\theta_1}{(\theta_1+\theta_2)}M_{x-1}(a+1,\theta_1+\theta_2,\theta_2) \tag{2.5}$$

for all integral values of $x>0$ with the initial condition

$M_0(a,\theta_1,0)=(1+\theta_1)^{-a}$ and $M_x(a,\theta_1,0)=\dfrac{(a+x-1)!}{(a-1)!\;x!}\dfrac{\theta_1^x}{(1+\theta_1)^{a+x}}$ for $x>0$, then the probability model $M_x(a,\theta_1,\theta_2)$ is the QNBD model $P_x(a,\theta_1,\theta_2)$.

***Proof.*** Integrating equation (2.4) w.r. to $\theta_2$ and by making use of the initial condition $M_0(a,\theta_1,0)=(1+\theta_1)^{-a}$, the constant $C_0=(1+\theta_1)^{-a}$ and thus

$$M_0(a,\theta_1,\theta_2)=(1+\theta_1)^{-a}=P_0(a,\theta_1,\theta_2)$$

For $x=1$, the difference differential equation (2.5) becomes

$$\frac{\partial}{\partial\theta_2}M_1(a,\theta_1,\theta_2)=-\frac{(a+1)}{(1+\theta_1+\theta_2)}M_1(a,\theta_1,\theta_2)$$

The integrating factor for the above is $(1+\theta_1+\theta_2)^{a+1}$ and the general solution is

$$M_1(a,\theta_1,\theta_2)=(1+\theta_1+\theta_2)^{-a-1}C_1$$

By making use of the initial condition $M_1(a,\theta_1,0)=a\theta_1(1+\theta_1)^{-a-1}$ we get $C_1=a\theta_1$ and

$$M_1(a,\theta_1,\theta_2)=a\theta_1(1+\theta_1+\theta_2)^{-a-1}=P_1(a,\theta_1,\theta_2)$$

Taking $x=2$ in (2.5) and making use of the result above, we get

$$\frac{\partial}{\partial\theta_2}M_2(a,\theta_1,\theta_2)=-\frac{2(a+2)}{(1+\theta_1+2\theta_2)}M_2(a,\theta_1,\theta_2)+\frac{a(a+1)\theta_1}{(1+\theta_1+2\theta_2)^{a+2}} \tag{2.6}$$

On multiplying the differential equation above with its integrating factor $(1+\theta_1+2\theta_2)^{a+2}$ and then integrating w.r.t. to $\theta_2$, we get

$$(1+\theta_1+2\theta_2)^{a+2} M_2(a,\theta_1,\theta_2) = a(a+1)\theta_1\theta_2 + C_2$$

By making use of the initial condition $M_2(a,\theta_1,0) = \frac{a(a+1)}{2!} \frac{\theta_1^2}{(1+\theta_1)^{a+2}}$, the constant

$C_2 = \frac{a(a+1)}{2!}\theta_1^2$. Therefore,

$$M_2(a,\theta_1,\theta_2) = \frac{a(a+1)}{2!} \frac{\theta_1(\theta_1+2\theta_2)}{(1+\theta_1+2\theta_2)^{a+2}} = P_2(a,\theta_1,\theta_2)$$

Using the result above in (2.5) for $x=3$, we get

$$\frac{\partial}{\partial\theta_2} M_3(a,\theta_1,\theta_2) = -\frac{3(a+3)}{(1+\theta_1+3\theta_2)} M_3(a,\theta_1,\theta_2) + \frac{a(a+1)(a+2)\theta_1(\theta_1+3\theta_2)}{(1+\theta_1+3\theta_2)^{a+3}}$$

The integrating factor for the differential equation above is $(1+\theta_1+3\theta_2)^{a+3}$ and the general solution is

$$(1+\theta_1+3\theta_2)^{a+3} M_3(a,\theta_1,\theta_2) = a(a+1)(a+2)\theta_1\theta_2\left[\theta_1 + \frac{3}{2}\theta_2\right] + C_3$$

By the initial condition $M_3(a,\theta_1,0) = \frac{a(a+1)(a+2)}{3!} \frac{\theta_1^3}{(1+\theta_1)^{a+3}}$, the constant

$C_3 = \frac{a(a+1)(a+3)}{3!}\theta_1^3$ and thus

$$M_3(a,\theta_1,\theta_2) = \frac{a(a+1)(a+2)}{3!} \frac{\theta_1(\theta_1+3\theta_2)^2}{(1+\theta_1+3\theta_2)^{a+3}} = P_3(a,\theta_1,\theta_2)$$

In a similar manner it can be easily shown by the method of induction that the unknown constants are determined by $C_x = \frac{a(a+1)....(a+x-1)}{x!}\theta_1^x$ for $x = 1,2,.....$

and that

$$M_x(a,\theta_1,\theta_2) = \frac{(a+x-1)!}{(a-1)!\,x!} \frac{\theta_1(\theta_1+x\theta_2)^{x-1}}{(1+\theta_1+x\theta_2)^{a+x}} = P_x(a,\theta_1,\theta_2)$$

for all nonnegative integral values of $x$.

## 3. Characterization Based on Conditional Distribution and by Damage Process

***Theorem.3.1.*** If $X$ and $Y$ are two independent random variables defined on the set of all non-negative integers such that

$$P\left[X=k\Big/X+Y=n\right]=\frac{\binom{n_1+k-1}{k}\binom{n_2+n-k-1}{n-k}}{\binom{n_1+n_2+n-1}{n}}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}(\theta_1+(n-k)\theta_2)^{n-k-1}(1+\theta_1+n\theta_2)^{n_1+n_2+n}}{(1+\theta_1+k\theta_2)^{n_1+k}(1+\theta_1+(n-k)\theta_2)^{n_2+k}(\theta_1+n\theta_2)^{n-1}} \tag{3.1}$$

for $k=0,1,2,.....n$ and zero else where

Then show that $X$ and $Y$ must have QNBD with parameters $(n_1,\theta_1,\theta_2)$ and $(n_2,\theta_1,\theta_2)$ respectively.

***Proof.*** Let $P(X=x)=f(x)>0$, $\sum f(x)=1$ and $P(Y=y)=g(y)>0$, $\sum g(y)=1$

By the given condition the random variables $X$ and $Y$ are independent, therefore

$$P\left[X=k\Big/X+Y=n\right]=\frac{f(k)g(n-k)}{\sum_{k=0}^{n}f(k)g(n-k)} \tag{3.2}$$

By making use of (3.1) in the equation above, for $n\geq 1$ and $0\leq k\leq n$, we get functional relation

$$\frac{f(k)g(n-k)}{f(k-1)g(n-k+1)}=\frac{(n_1+k-1)(n-k+1)}{k(n_2+n-k)}\frac{(\theta_1+k\theta_2)^{k-1}}{(\theta_1+(k-1)\theta_2)^{k-2}}\frac{(\theta_1+(n-k)\theta_2)^{n-k-1}}{(\theta_1+(n-k+1)\theta_2)^{n-k}}$$

$$\times\frac{(1+\theta_1+(k-1)\theta_2)^{n_1+k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}\frac{(1+\theta_1+(n-k+1)\theta_2)^{n_2+n-k+1}}{(1+\theta_1+(n-k)\theta_2)^{n_2+n-k}} \tag{3.3}$$

Replacing $k$ by $k+1$ and $n$ by $n+1$ in the above, we get

$$\frac{f(k+1)g(n-k)}{f(k)g(n-k+1)}=\frac{(n_1+k)(n-k+1)(\theta_1+(k+1)\theta_2)^{k}}{(k+1)(n_2+n-k)(\theta_1+k\theta_2)^{k-1}}\frac{(\theta_1+(n-k)\theta_2)^{n-k-1}}{(\theta_1+(n-k+1)\theta_2)^{n-k}}$$

$$\times\frac{(1+\theta_1+k\theta_2)^{n_1+k}}{(1+\theta_1+(k+1)\theta_2)^{n_1+k+1}}\frac{(1+\theta_1+(n-k+1)\theta_2)^{n_2+n-k+1}}{(1+\theta_1+(n-k)\theta_2)^{n_2+n-k}} \tag{3.4}$$

On dividing (3.4) by (3.3), we get

$$\frac{f(k+1)f(k-1)}{[f(k)]^2}=\frac{(n_1+k)k}{(k+1)(n_1+k-1)}\frac{(\theta_1+(k+1)\theta_2)^{k}(\theta_1+(k-1)\theta_2)^{k-2}}{(\theta_1+k\theta_2)^{2k-2}}$$

$$\times\frac{(1+\theta_1+k\theta_2)^{2n_1+2k}}{(1+\theta_1+(k+1)\theta_2)^{n_1+k+1}(1+\theta_1+(k-1)\theta_2)^{n_1+k-1}}$$

Taking $k=1,2,.....n-1$ in the equation above and then on multiplying together, we get

$$\frac{f(n)}{f(n-1)}=\frac{f(1)}{f(0)}\frac{(n_1+n-1)}{nn_1}\frac{(\theta_1+n\theta_2)^{n-1}}{\theta_1(\theta_1+(n-1)\theta_2)^{n-2}}\frac{(1+\theta_1+\theta_2)^{n_1+1}}{(1+\theta_1)^{n_1}}\frac{(1+\theta_1+(n-1)\theta_2)^{n_1+n-1}}{(1+\theta_1+n\theta_2)^{n_1+n}}$$

Setting $\frac{f(1)}{f(0)}=\frac{n_1\theta_1(1+\theta_1)^{n_1}}{(1+\theta_1+\theta_2)^{n_1+1}}$ in the equation above, we get

$$f(n)=\frac{(n_1+n-1)}{n}\frac{(\theta_1+n\theta_2)^{n-1}}{(\theta_1+(n-1)\theta_2)^{n-2}}\frac{(1+\theta_1+(n-1)\theta_2)^{n_1+n-1}}{(1+\theta_1+n\theta_2)^{n_1+n}}f(n-1)$$

A repeated use of the equation above gives

$$f(n)=\frac{(n_1+n-1)!}{(n_1-1)!\,n!}\frac{\theta_1(\theta_1+n\theta_2)^{n-1}(1+\theta_1)^{n_1}}{(1+\theta_1+n\theta_2)^{n_1+n}}f(0)$$

Now, by making use of the fact that $\sum f(x)=1$, the recurrence relation above gives $f(0)=(1+\theta_1)^{-n_1}$ and thus

$$f(x)=\frac{(n_1+x-1)!}{(n_1-1)!\,x!}\frac{\theta_1(\theta_1+x\theta_2)^{x-1}}{(1+\theta_1+x\theta_2)^{n_1+x}} \qquad x=0,1,2,\ldots..$$

Hence the random variable $X$ has a QNBD with parameters $(n_1,\theta_1,\theta_2)$.

On taking $k=1$ in (3.3), we get

$$\frac{g(n)}{g(n-1)}=\frac{f(0)}{f(1)}\frac{(n_2+n-1)}{nn_1}\frac{(\theta_1+n\theta_2)^{n-1}}{\theta_1(\theta_1+(n-1)\theta_2)^{n-2}}\frac{(1+\theta_1+\theta_2)^{n_1+1}}{(1+\theta_1)^{n_1}}\frac{(1+\theta_1+(n-1)\theta_2)^{n_2+n-1}}{(1+\theta_1+n\theta_2)^{n_2+n}}$$

Setting as usual $\frac{f(1)}{f(0)}=\frac{n_1\theta_1(1+\theta_1)^{n_1}}{(1+\theta_1+\theta_2)^{n_1+1}}$ in the equation above, we get

$$g(n)=\frac{(n_2+n-1)}{n}\frac{(\theta_1+n\theta_2)^{n-1}}{(\theta_1+(n-1)\theta_2)^{n-2}}\frac{(1+\theta_1+(n-1)\theta_2)^{n_2+n-1}}{(1+\theta_1+n\theta_2)^{n_2+n}}g(n-1)$$

A repeated use of the equation above gives

$$g(n)=\frac{(n_2+n-1)!}{(n_2-1)!\,n!}\frac{\theta_1(\theta_1+n\theta_2)^{n-1}(1+\theta_1)^{n_2}}{(1+\theta_1+n\theta_2)^{n_2+n}}g(0)$$

Since $\sum g(x)=1$, the recurrence relation above gives $g(0)=(1+\theta_1)^{-n_2}$ and thus

$$g(y)=P(Y=y)=\frac{(n_2+y-1)!}{(n_2-1)!\,y!}\frac{\theta_1(\theta_1+y\theta_2)^{y-1}}{(1+\theta_1+y\theta_2)^{n_2+y}} \qquad y=0,1,2,\ldots..$$

Hence the random variable $Y$ also possesses QNBD with parameters $(n_2,\theta_1,\theta_2)$.

***Theorem. 3.2.*** Let $X_1$ and $X_2$ be two independent discrete random variables whose sum $Y$ is a QNB variate with parameters $(a,\theta_1,\theta_2)$. Then $X_1$ and $X_2$ must each be a QNB variate defined over all non-negative integers.

***Proof.*** Consul (1974) and famoye (1994) proved this theorem for GPD and GNBD respectively. Since QNBD is a generalization, in Gurland's (1957) terminology, of restricted GPD model with parameters $(\theta,\alpha\theta)$ obtained by compounding the GPD model through the values of $\theta$ with gamma distribution $\gamma(a,b)$ as mixing distribution. Further, QNBD is also generalization of NBD. Therefore, taking these facts into consideration, an attempt has been made here to prove that the theorem under consideration also holds good for the proposed model.

Since the QNB variate $Y$ has a lattice distribution defined over all non-negative integers. Therefore, using arguments of Raikov (1937), the random variables $X_1$ and $X_2$ must have also lattice distribution defined over all non-negative integers.

Let the pgf of the random variable $X_i, i=1,2$ be denoted by

$$g_i(u)=\sum_{x=0}^{\infty}P_i(x)u^x, \qquad |U|<1$$

Where $P_i(x)=P(X_i=x)$ represents the pdf of $X_i, i=1,2$ .Since the sum $Y=X_1+X_2$ has a QNBD with parameters $(a,\theta_1,\theta_2)$, therefore its pgf is

$g(u)=\phi(t)=(1+\theta_1-\theta_1 t)^{-a}$ where $t=(1+\theta_2-\theta_2 t)^{-a}$ and

$g_1(u)g_2(u)=g(u)=\phi(t)=(1+\theta_1-\theta_1 t)^{-a}$ where $t=(1+\theta_2-\theta_2 t)^{-a}$

Now, using the arguments of Raikov (1937), the pgf of QNBD can only be factorized into pgf's of negative binomial distributions. Thus, the factors $\phi_1(t)$ and $\phi_2(t)$ of $\phi(t)=(1+\theta_1-\theta_1 t)^{-a}$ must be given by $\phi_1(t)=(1+\theta_1-\theta_1 t)^{-a_1}$ and $\phi_2(t)=(1+\theta_1-\theta_1 t)^{-a+a_1}$ where $a_1$ is an arbitrary number such that $0<a_1<a$. Hence pgf's of $X_1$ and $X_2$ becomes

$g_1(u)=(1+\theta_1-\theta_1 t)^{-a_1}$ and $g_2(u)=(1+\theta_1-\theta_1 t)^{-a+a_1}$ where $t=(1+\theta_2-\theta_2 t)^{-a}$. Because of the uniqueness property, the pgf's $g_1(u)$ and $g_2(u)$ must represent QNBD models. Thus, $X_1$ and $X_2$ must be a QNBD variate defined over all non-negative integers with parameters $(a_1,\theta_1,\theta_2)$ and $(a-a_1,\theta_1,\theta_2)$ respectively, where $0<a_1<a$.

***Theorem. 3.3.*** If a non-negative QNB variate $X$ is subdivided into two components $X_1$ and $X_2$ such that the conditional distribution $P\left[X_1=k, X_2=x-k \big/ X=x\right]$ is a hypergeometric-quazi-negative-binomial distribution

$$\frac{\binom{a_1+k-1}{k}\binom{a-a_1+x-k-1}{x-k}}{\binom{a+x-1}{x}}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}(\theta_1+(x-k)\theta_2)^{x-k-1}(1+\theta_1+x\theta_2)^{a+x}}{(1+\theta_1+k\theta_2)^{a_1+k}(1+\theta_1+(x-k)\theta_2)^{a-a_1+x-k}(\theta_1+x\theta_2)^{x-1}} \quad \text{with}$$

Parameters $(a,a_1,k,\theta_1,\theta_2)$, $0<a_1<a$ then the random variables $X_1$ and $X_2$ are independent and have the QNBD.

***Proof.*** Let $X$ be a QNB variate with parameters $(a,\theta_1,\theta_2)$ then its probability distribution is

$$P(X=x)=\binom{a+x-1}{x}\frac{\theta_1(\theta_1+\theta_2 x)^{x-1}}{(1+\theta_1+\theta_2 x)^{a+x}} \qquad x=0,1,2,.....$$

The joint probability distribution of random variables $X_1$ and $X_2$ is given by the conditional distribution

$$P(X_1=k, X_2=x-k)=P\left[X_1=k, X_2=x-k \big/ X=x\right]P(X=x)$$

On substituting the values in the equation above, we get

$$P(X_1=k, X_2=x-k)=\binom{a_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{a+k}}\binom{a-a_1+x-k-1}{x-k}\frac{\theta_1(\theta_1+(x-k)\theta_2)^{x-k-1}}{(1+\theta_1+\theta_2 x)^{a-a_1+x-k}}$$

$$=P_k(a,\theta_1,\theta_2)\,P_{x-k}(a-a_1,\theta_1,\theta_2)$$

Which is a product of two quazi-negative-binomial probabilities corresponding to two random variables $X_1$ and $X_2$. Thus the two random variables are independent and have QNBD models with parameters $(a,\theta_1,\theta_2)$ and $(a-a_1,\theta_1,\theta_2)$ respectively.

***Characterization by damage process.***

When an investigator collects a sample of observations produced by nature, according to a certain model, the original distribution may not be produced due to non-observability of some events or the partial destruction of some units. This problem was first pointed out by Rao (1963) when he studied the resultant model after the observations, generated by some probability models, were damaged by other probability models. Subsequently Rao and Rubin (1964), Srivastava and Srivastava (1970) and Consul (1975) also studied the same problem in case of Poisson and GPD models. In this paper, we have made an attempt to obtain some characterization of QNBD model in the light of the same theory.

Let $X$ be a random variable defined on non-negative integers with probability distribution $\{P_n\}$ and let $Z$ be a random variable denoting the undamaged part of the random variable $X$ when it is subject to destruction process such that $P\left[Z=k/_{X=x}\right]=S\left[k/_x\right]$, $k=0,1,\ldots\ldots x$, then we have the following theorem.

***Theorem. 3.4.*** If $X$ is a QNB variate with parameters $(n,\theta_1,\theta_2)$ and if the destructive process is hypergeometric-quazi-negative-binomial variate given by

$$S\left[k/_x\right]=\frac{\binom{n_1+k-1}{k}\binom{n-n_1+x-k-1}{x-k}}{\binom{n+x-1}{x}}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}(\theta_1+(x-k)\theta_2)^{x-k-1}(1+\theta_1+x\theta_2)^{n+x}}{(1+\theta_1+k\theta_2)^{n_1+k}(1+\theta_1+(x-k)\theta_2)^{n-n_1+x-k}(\theta_1+x\theta_2)^{x-1}}$$

$$k=0,1,\ldots\ldots x$$

Show that

i) $Z$ is a QNB variate with parameters $(n_1,\theta_1,\theta_2)$.

ii) $P(Z=k)=P\left[Z=k/_{X\,damaged}\right]=P\left[Z=k/_{X\,undamaged}\right]$

***Proof.***

i) $$P(Z=k)=\sum_{x=k}^{\infty}S\left[k/_x\right]P_x(n,\theta_1,\theta_2)$$

$$=\sum_{x=k}^{\infty}\binom{n_1+k-1}{k}\binom{n-n_1+x-k-1}{x-k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}(\theta_1+(x-k)\theta_2)^{x-k-1}\,\theta_1}{(1+\theta_1+k\theta_2)^{n_1+k}(1+\theta_1+(x-k)\theta_2)^{n-n_1+x-k}}$$

$$=\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}\sum_{r=0}^{\infty}\binom{n-n_1+r-1}{r}\frac{(\theta_1+r\theta_2)^{r-1}\,\theta_1}{(1+\theta_1+r\theta_2)^{n-n_1+r}}$$

$$=\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}=P_k(n_1,\theta_1,\theta_2)$$

Thus the random variable $Z$ is a QNB variate with parameters $(n_1,\theta_1,\theta_2)$.

ii) $$P\left[Z=k/_{X\,damaged}\right]=\frac{\sum_{x=k}^{\infty}S\left[k/_x\right]P_x(n,\theta_1,\theta_2)}{\sum_{k=0}^{\infty}\sum_{x=k}^{\infty}S\left[k/_x\right]P_x(n,\theta_1,\theta_2)}$$

$$=\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}\div\left[\sum_{k=0}^{\infty}\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}\right]$$

$$=\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}=P(Z=k)$$

Similarly it can be shown that

$$P\left[Z=k/X\ undamaged\right]=\frac{S\left[k/k\right]P_k(n,\theta_1,\theta_2)}{\sum_{k=0}^{\infty}S\left[k/k\right]P_k(n,\theta_1,\theta_2)}$$

$$=\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}=P(Z=k)$$

where $\sum_{k=0}^{\infty}S\left[k/k\right]P_k(n,\theta_1,\theta_2)=\sum_{k=0}^{\infty}\binom{n_1+k-1}{k}\frac{\theta_1(\theta_1+k\theta_2)^{k-1}}{(1+\theta_1+k\theta_2)^{n_1+k}}=1$

## 4. QNBD Model as a Distribution of Number of Accidents in the Light of Lrwin's Theory of "Proneness-Liability" Model

Traffic accidents remain in concern for every one's life. Various theories have been developed concerning the interpretation of different situations. A natural model which assumes that the probabilities of having an accident are only result of random factors is that the number of accidents is poisson distribution with parameter $\lambda$ i.e,

$$P(N=n)=\frac{\lambda^n}{n!}e^{-\lambda}, \qquad n=0,1,2,........$$

Where $N$ is the random variable which describes the number of accidents of a single person.

Another theory, the "accident proneness" theory which takes into account the indual's difference in probabilities of having an accident or in their "accident proneness" which remain constant in time. This theory takes into consideration both the factors random as well as non-random where the non-random factors refers to indual's psychology, explaining in this way, more or less, why the indual's have unequal accident proneness. Taking these situations into consideration, Greenwood and Woods (1919), obtained that the number of accidents $N$ has a negative binomial distribution with parameter $k$ and $\frac{1}{v}$ i.e.,

$$P(N=n)=\binom{k+n-1}{n}v^n(1+v)^{n+k} \qquad n=0,1,2,........$$

Consul (1989) also described the use of GPD model in an accident theory and applied it to a number of data sets pertaining to Shunting accidents, home injuries, and strikes in industries and obtained a better fit as compared to poisson and negative binomial distribution. The GPD model is

$$P(N=n)=\frac{\lambda(\lambda+n\theta)^{n-1}e^{-(\lambda+n\theta)}}{n!}, \qquad n=0,1,2,........$$

He also gave the interpretation of the parameters-the parameter $\lambda$ represents the "accident proneness" and $\theta$ represents the "rate of restitution process" of the subject under study.

The Irwin's theory of "proneness-liability" model which assumes also that the non-random factors can be further split into psychological and external factors provides more explanation as to why some indual's in the population tend to have more accidents than others. In the context of this model, the indual accident proneness does not remain constant, because the population is exposed to a variable risk. In his model, Irwin used the term "accident proneness" to refer to a person's predisposition to accidents, and the term "accident-liability" to refer to a person's exposure to external risk of the accident and he derived the univariate generalized Waring distribution as the distribution of number of accidents. Irwin (1975b) applied this model to data on accidents sustained by men in a soap factory, providing an improved fit as compared to the negative binomial.

In fact, Irwin derived his model by compounding the parameter of Poisson distribution with gamma distribution thus resulting in NBD which when compounded with beta-$II$ distribution gives Irwin's model. This model is one interesting member of the family of mixed Poisson distribution. Making this as basis, here, we have made an attempt to derive the distribution of the number of accidents $N$ in the light of Irwin's theory starting with the restricted Consul and Jain's (1973) GPD model

$$P(N=n)=\frac{\lambda^n(1+n\alpha)^{n-1}e^{-\lambda(1+n\alpha)}}{n!}, \qquad n=0,1,2,........ \tag{4.1}$$

As per the interpretation of the Consul (1989), the parameter $\lambda$ represents the "accident proneness" and $\theta=\alpha\lambda$ represents the "rate of restitution process" of the subject under study. In the light of Irwin's theory, the indual accident proneness does not remain constant in time, and thus if we allow the parameter $\lambda$ in the model (4.1) to follow gamma distribution with parameters $k$ and $\frac{1}{v}$ i.e. $\lambda$ for given $v$ is a random variable with density given by

$$f(\lambda/v)=\frac{v^{-k}}{\Gamma(k)}\lambda^{k-1}e^{-\lambda v^{-1}} \qquad \lambda\geq 0 \tag{4.2}$$

and then on compounding (4.1) through the values of $\lambda$ by (4.2), we get the distribution of the number of accidents $N$ as the quasi-negative binomial distribution

$$P(N=n)=\binom{k+n-1}{n}\frac{v^{-k}(1+n\alpha)^{n-1}}{(1+v^{-1}+n\alpha)^{k+n}} \qquad n=0,1,2,........$$

Taking $\theta=\alpha v$, the distribution becomes

$$P(N=n)=\binom{k+n-1}{n}\frac{v(v+n\theta)^{n-1}}{(1+v+n\theta)^{k+n}} \qquad n=0,1,2,........ \tag{4.3}$$

Where $(k,v,\theta)$ represents the parameters of the QNBD.

Hence, in the light of Irwin's theory of "proneness-liability" model, the QNB model explains both the variations in accident-proneness as well as in accident-liabilities of the subject under consideration..

## 5. Application of QNBD Model in Shunting Accidents, Home Injuries, and Strikes in Industries

In this section, we are more interested in the fitting of the proposed model and its comparison with the GPD model but not in the application of the chi-square test for testing the significance of the discrepancies between the observed and expected frequencies as the degree of freedom provided by most of the data sets (tables (5.1)-(5.4)) for the proposed model is Zero which makes the chi-square test of significance invalid.

. The data sets in tables (5.1) to (5.3) were previously used by Adelstein (1952) for Poisson and negative binomial distributions and concluded that the negative binomial fits well than Poisson distribution. Consul (1989a) used the same data sets for GPD model and reached to a conclusion that GPD model gives best fit than Poisson and negative binomial distributions, for more explanation and details; see GPD Consul (1989), pages 117-121. Here, we applied the proposed model to the same data sets. The parameters of the proposed model have been estimated by ML method with the help of a computer programme in R-soft wear.

TABLE 5.1

Comparison of observed frequencies for first-year shunting accidents and for a five year record of experienced men with expected QNBD frequencies for different age groups.

| No.of accidents | Age 21-25 yr | | Age 26-30 yr | | Age 31-35 yr | | 5-yr record for experience | |
|---|---|---|---|---|---|---|---|---|
| | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD |
| 0 | 80 | 76.30 | 121 | 123.39 | 80 | 80.29 | 54 | 51.20 |
| 1 | 56 | 65.15 | 85 | 80.18 | 61 | 60.28 | 60 | 62.74 |
| 2 | 30 | 23.50 | 19 | 20.66 | 13 | 13.56 | 36 | 40.51 |
| 3 | 4 } | 5.10 | 1 | 2.61 | 1 | 0.86 | 21 | 18.35 |
| $\geq 4$ | 0 } | | 1 | 0.16 | 0 | 0.01 | 11 | 9.26 |
| Total | 170 | 170 | 227 | 227 | 1 55 | 155 | 182 | 182 |
| ML Estimate | $a=21.94218080$<br>$\theta_1=0.037215153$<br>$\theta_2=-0.00369338$ | | $a=21.942168064$<br>$\theta_1=0.028172153$<br>$\theta_2=-0.00347625$ | | $a=38.1442428$<br>$\theta_1=0.0173925$<br>$\theta_2=-.00365469$ | | $a=41.94190$<br>$\theta_1=0.03070$<br>$\theta_2=0.00046$ | |
| | 3.504626 | | 0.683468 | | 0.05219945 | | 1.48454 | |

TABLE 5.2

Comparison of observed frequencies of accidents of 122 experienced shunting men over 11 years (1937-1947) with expected QNBD frequencies

| No.of accidents | 1937-1942 | | 1943-1947 | | 1937-1947 | |
|---|---|---|---|---|---|---|
| | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD |
| 0 | 40 | 39.86 | 50 | 50.88 | 21 | 20.07 |
| 1 | 39 | 39.68 | 43 | 40.52 | 31 | 30.55 |
| 2 | 26 | 23.80 | 17 | 19.51 | 26 | 27.53 |
| 3 | 8 | 11.24 | 9 | 7.46 | 19 | 19.29 |
| 4 | 6 | 4.65 | 2 } | | 7 | 11.66 |
| 5 | 2 } | | 0 } | 3.63 | 9 | 6.43 |
| 6 | 1 } | 2.77 | 1 } | | 9 | 6.47 |
| Total | 122 | 122 | 122 | 122 | 122 | 122 |
| ML Estimate | $a=21.942214414$ | | $a=11.209125667$ | | $a=31.942233844$ | |
| | $\theta_1=$ 0.05230512 | | $\theta_1=$ 0.081133953 | | $\theta_1=0.058137631$ | |
| | $\theta_2=$ 0.00419416 | | $\theta_2=$ 0.004864953 | | $\theta_2=0.004578100$ | |
| $\chi^2$ | 1.560489 | | 0.917171 | | 4.018035 | |

TABLE 5.3

Comparison of observed frequencies for home injuries of 122 experienced men during 11 years (1937-1947) with the expected QNBD frequencies

| No.of Injuries | 1937-1942 | | 1943-1947 | | 1937-1947 | |
|---|---|---|---|---|---|---|
| | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD |
| 0 | 73 | 73.23 | 88 | 87.92 | 58 | 57.07 |
| 1 | 36 | 35.34 | 18 | 18.78 | 34 | 34.68 |
| 2 | 10 | 10.3 | 11 | 9.71 | 14 | 16.63 |
| 3 | 2 } | | 4 } | | 8 | 7.50 |
| 4 | 1 } | 3.05 | 1 } | 5.59 | 6 } | |
| 5 | - | - | - | - | 2 } | 0.12 |
| Total | 122 | 122 | 122 | 122 | 122 | 122 |
| ML Estimate | $a=6.929456168$ | | $a=0.2276245$ | | $a=32.942033133$ | |
| | $\theta_1=$ 0.076422976 | | $\theta_1=$ 3.2170325 | | $\theta_1=0.023331560$ | |
| | $\theta_2=0.002652596$ | | $\theta_2=-0.6553422$ | | $\theta_2=0.006401901$ | |
| $\chi^2$ | 0.0277794 | | 0.2661209 | | 1.055267 | |

After close examination of the data sets in tables (5.1) to (5.3) we conclude that the proposed model fits well than GPD model except that the aggregated data in the last columns of table (5.1) and table (5.2). As explained in section 4 of this paper, the best fit obtained is due to the fact that the proposed model explains both the variation in accident-proneness as well as in accident-liabilities of the subject under study.

Now, we present more data sets in table (5.4) on the number of strikes in 4-week periods in four leading industries in the united Kingdom during (1948-1959) were previously used by Kendall (1961) and concluded that the aggregate data for the four industries agrees with the poisson law but that it does not hold that well for the indual industries. Consul (1989) used the same data sets for GPD model and observed that the data follows GPD model in Vehicle manufacturing industry, Ship-building industry, and Transport industry but that the pattern in the Coal-mining industry can not be well described by GPD model. We also applied the proposed model to the same data sets and the expected frequencies are shown in table (5.4), the parameters have been estimated by ML method with the help of a computer programme in R-soft wear.

TABLE 5.4

Comparison of observed frequencies of the number of outbreaks of strike in four leading industries in the U.K. during (1948-1959) with the expected QNBD frequencies

| No.of outbreaks | Coal mining | | Vehicle manufacture | | Ship building | | Transport | |
|---|---|---|---|---|---|---|---|---|
| | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD | Obs. | QNBD |
| 0 | 46 | 50.22 | 110 | 109.79 | 117 | 116.73 | 114 | 114.84 |
| 1 | 76 | 65.42 | 33 | 33.45 | 29 | 30.27 | 35 | 31.90 |
| 2 | 24 | 32.29 | 9 | 9.19 | 9 | 6.94 | 4 | 7.23 |
| 3 | 9 } | 8.07 | 3 | 3.57 | 0 | 2.06 | 2 | 2.03 |
| ≥ 4 | 1 } | | 1 | | 1 | | 1 | |
| Total | 156 | 156 | 156 | 156 | 156 | 156 | 156 | 156 |
| ML Estimate | $a = 32.94235746$ $\theta_1 = 0.035003245$ $\theta_2 = -0.00475275$ | | $a = 37.94234634$ $\theta_1 = 0.009301366$ $\theta_2 = 0.003574957$ | | $a = 38.942347767$ $\theta_1 = 0.007475730$ $\theta_2 = 0.00273226$ | | $a = 32.09423493$ $\theta_1 = 0.009588379$ $\theta_2 = 0.00283426$ | |
| | 4.655564 | | 0.06217639 | | 1.210815 | | 2.213897 | |

After comparing the chi-square values of the proposed model with the GPD model we did not find any improvement in fitting by the proposed model for the table (5.4) and observed almost equal fit.

## REFERENCE

Greenwood, M. and Woods, II. M. (1919). On the incidents of the Industrial Accidents upon Individuals with special reference to Multiple Accidents. Report of the Industrial Fatigue Rearch Board. 4, 1-28. London: His Majesty's Stationery Office.

Raikov, D. (1937). A characteristic property of the poisson laws, C. R. Acad. Sci. U.S.S.R., 14, 8-11.

Tukey, John W. (1949). Moments of random group size distributions, Ann. Math. Statist. 20, 523-539.

Adelstein, A.M. (1952). Accident proneness: a criticism of the concept based upon an analysis of shunter's accidents, J. Roy. Statist. Soc. Ser. A, 115,354-410.

Gurland, J. (1957). Some interrelations among compound and generalized distributions. Biometrika 44, 265-68.

Kendal, M. G. (1961). Natural law in the social sciences, J. Roy. Statist. Soc. Ser. A, 124, 1-19.

Rao, C. R. (1963). On discrete distributions arising out of the methods of ascertainment. Int. Symp. Discrete distributions, Monetreal.

Rao, C. R. and Robin, H. (1964). On a characterization of the Poisson distribution. Sankhya, Ser. A, 26,295-298.

Srivastava, R. C. and Srivastava, A. B. L. (1970). On a characterization of the Poisson distribution. J. Appl. Prob., 7, 497-501.

Consul, P. C. and Jain, G.C. (1973): A generalization of the Poisson distribution, Techno-metrics, 15(4), 791-799.

Consul, P. C. (1974). On a characterization of Lagrangian Poisson and quasi binomial distributions. Comm. Statist., A- Theory Methods, 4(6), 555-563.

Janardan, K. G. (1975): Markov-Polya urn-model with pre-determined strategies, Gujarat Statist. Rev., 2 (1), 17-32.

Consul, P. C. (1975). Some new characterization of discrete Lagrangian distributions. In statistical distributions in scientific work, Vol-3, 279-290: Eds. G. P. Patil, S. Kotz, and J. Ord. D. Reidel, Boston.

Irwin, J. I. (1975b). "The Generalized Waring Distribution. Parts II." Journal of the Royal Statistical Society, A, 138, 204-227.

Consul, P. C. (1988). On Some models leading to the generalized Poisson distribution. Comm. Statist., A- Theory Methods, 17,423-442.

Consul, P. C. (1989). Generalized Poisson distributions properties and applications, Marcel Dekker, Inc., New York.

Consul, P. C. (1990d). On some properties and applications of quasi-binomial distribution. Comm. statist. - Theory Methods, 19, 477-504.

S.B.Nand and K. K. Das (1994): A family of Abel series distribution, Sankhya, B, 56(2), 147-164.

Famoye, F. (1994). Characterization of generalized negative binomial distribution. Journal of mathematical sciences, 5, 71-81.

Sen, K. and Jain, R. (1996): Generalized Markov-Polya urn-model with pre-determined strategies, J. Statist. Plann. Infer. 54, 119-133.