Applied Mathematical Sciences, Vol. 7, 2013, no. 22, 1093 - 1105
HIKARI Ltd, www.m-hikari.com

The Negative Binomial-Generalized Exponential (NB-GE) Distribution

Sirinapa Aryuyuen

Department of Statistics, Faculty of Science
Kasetsart University, Bangkok 10900, Thailand
g5517400469@ku.ac.th

Winai Bodhisuwan[1]

Department of Statistics, Faculty of Science
Kasetsart University, Bangkok 10900, Thailand
fsciwnb@ku.ac.th

Abstract

This paper introduced a new mixed distribution, namely the negative binomial-generalized exponential (NB-GE) distribution, which it obtained by mixing the negative binomial (NB) distribution with a generalized exponential (GE) distribution. The NB-GE distribution offers the advantage of being able to handle this kind of data sets, while still maintaining similar characteristics as the traditional negative binomial. The closed form and the factorial moment of the NB-GE distribution are derived. The NB-GE distribution appropriate when we find the one related to count data which have a large number of zeros and overdispersion in Poisson distribution. In addition, we present the basic properties of the new distribution such as mean, variance, skewness and kurtosis. Including, the negative binomial-exponential distributions are presented as special cases of this NB-GE distribution. Parameters estimation is also implemented using maximum likelihood method and the usefulness of the NB-GE distribution is illustrated by real data set. When using a goodness of fit test, the results show that the NB-GE distribution can provide a better fit than the NB distribution and Poisson distribution for count data that contain a large number of zeros.

Mathematics Subject Classification: 60E05

[1]Corresponding Author

Keywords: a large number of zero, mixed negative binomial distribution, negative binomial-generalized exponential distributions, overdispersion

1 Introduction

The Poisson distribution plays an important role in the modeling process even if it is not applicable, because it is often used to derive more realistic model that meet characteristics of observed data to be used explain parameter of distribution. The Poisson distribution is usually handle to fit the count data in practice when number of phenomenon are randomly distributed over time and/or space the counts of the phenomenon occurring. Equality of mean and variance are characteristic of the Poisson distribution, but in practice, however, the observed count data often displays features like overdispersion (variance is greater than the mean) or underdispersion (variance is smaller than the mean), which is common in applied data analysis [11].

One frequent manifestation of over-dispersion is that the incidence of zero counts is greater than expected for the Poisson distribution and this is of interest because zero counts frequently have special status, which is a violation of the Poisson restriction that the variance of the observed random variable equal its mean [15]. Statistical procedures must eliminate for these sources of variability in count data, which suggested a model in which the mean of Poisson distribution has a gamma distribution. Practically, this leads to the negative binomial (NB) distribution.

The NB distribution has become increasingly popular as a more flexible alternative to Poisson distribution, especially when it is questionable whether the strict requirements for Poisson distribution could be satisfied [5]. But NB distribution is better for overdispersed count data that are not necessarily heavy-tailed. The extremely heavy tail implies overdispersion, but the converse does not hold [14]. Among the count data issues, which have a large number of zeros that lead to heavy tail. For such data sets, the number of sites where no crash is observed is so large that traditional statistical distributions or models, such as the Poisson and NB distributions, cannot be used efficiently. The Poisson distribution tends to underestimate the number of zeros given the mean of the data, while the NB distributions may overestimate zeros, but underestimate observations with a count [7].

The distribution mixtures define one of the most important ways to obtain new probability distributions in applied probability and operational research [3]. In this sense, we are looking for a more flexible alternative to the Poisson distribution, especially under the overdispersion, the NB distribution [13] obtained as a mixture of Poisson and gamma distributions, NB-exponential distribution [12], NB-Pareto distribution [8], NB-inverse Gaussian distributions [3], NB-Lindley distribution [16] and NB-beta exponential distribution

[10] have been proposed in the count data are either overdispersion.

This paper introduced a new mixed NB distribution, which its obtained by mixing a NB distribution with a generalized exponential (GE) distribution. The incurred distribution is called negative binomial-generalized exponential (NB-GE) distribution, which introduces two more free parameter and is more flexible in fitting count data. Including, we present properties of the NB-GE distribution are the factorial moments, the first four moments, variance, skewness, and kurtosis. The parameters of NB-GE distribution are estimated by maximum likelihood Estimation (MLE), and presents the comparison analysis between the Poisson, NB and NB-GE distributions using real data set.

2 Materials and Methods

2.1 The negative binomial-generalized exponential (NB-GE) distribution

The NB distribution is often employed in case where a distribution is overdispersed. If X denotes a random variable distributed under a NB distribution with parameter r and p, then its probability mass function (pmf) is given by:

$$f_1(x; r, p) = \binom{r+x-1}{x} p^r (1-p)^x, \tag{1}$$

where $x = 0, 1, 2, \ldots$ for $r > 0$ and $0 < p < 1$. It is consequently, we obtain:

$$E(X) = \frac{r(1-p)}{p} \quad \text{and} \quad E(X^2) = \frac{r(1-p)(1+r(1-p))}{p^2}.$$

The factorial moment of X is:

$$\mu_{[k]}(X) = E[X(X-1)\cdots(X-k+1)] = \frac{\Gamma(r+k)}{\Gamma(r)} \frac{(1-p)^k}{p^k},\ k = 1, 2, \ldots, \tag{2}$$

where $\Gamma(\cdot)$ is the gamma function defined by:

$$\Gamma(t) = \int_0^\infty x^{t-1} e^{-x} \mathrm{d}x,\ t > 0.$$

The form of a NB-GE distribution which is a mixed NB distribution obtained by mixing the distribution of $\mathrm{NB}(r, p)$ where, $p = \exp(-\lambda)$, $\lambda > 0$ with distribution of $\mathrm{GE}(\alpha, \beta)$. The GE distribution has probability density function (pdf):

$$f_2(x; \alpha, \beta) = \alpha\beta(1 - e^{\beta x})^{\alpha-1} e^{-\beta x},\ x > 0,\ \text{for } \alpha, \beta > 0. \tag{3}$$

The GE distribution introduced by Gupta and Kundu [4], and shown in there that its moment generating function (mgf) is given by:

$$M_X(t) = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{t}{\beta})}{\Gamma(\alpha-\frac{t}{\beta}+1)}. \tag{4}$$

As discussed above, the NB-GE distribution is a mixture of negative binomial and generalized exponential distribution, which has a heavy tail. We first provide a general definition of this distribution which will subsequently expose its probability mass function.

Definition 2.1 *Let X be a random variable of a NB-GE(r, α, β) distribution, $X \sim$ NB-GE(r, α, β), when the NB distribution have parameters $r > 0$ and $p = \exp(-\lambda)$, where λ is distributed as GE distribution with positive parameters α and β, i.e., $X|\lambda \sim$ NB$(r, p = \exp(-\lambda))$ and $\lambda \sim$ GE(α, β).*

Theorem 2.2 *Let $X \sim$ NB-GE(r, α, β). The probability mass function of X is given by:*

$$f(x; r, \alpha, \beta) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \left(\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right), \tag{5}$$

where $x = 0, 1, 2, \ldots$, r, α and $\beta > 0$.

Proof. If $X|\lambda \sim \text{NB}(r, p = \exp(-\lambda))$ in (1) and $\lambda \sim \text{GE}(\alpha, \beta)$ in (3), then the pmf of X can be obtained by:

$$f(x; r, \alpha, \beta) = \int_0^\infty f_1(x|\lambda) f_2(\lambda; \alpha, \beta) \mathrm{d}\lambda \tag{6}$$

where, $f_1(x|\lambda)$ is defined by:

$$f_1(x|\lambda) = \binom{r+x-1}{x} e^{-\lambda x}(1-e^{-\lambda})^x = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j e^{-\lambda(r+j)}. \tag{7}$$

By substituting (7) into (6), we obtain:

$$\begin{aligned} f(x|\lambda) &= \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \int_0^\infty e^{-\lambda(r+j)} f_2(\lambda; \alpha, \beta) \mathrm{d}\lambda \\ &= \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j M_\lambda(-(r+j)). \end{aligned} \tag{8}$$

Substituting the moment generating function of GE distribution (4) into (8), the pmf of NB-GE (r, α, β) is finally given as:

$$f(x; r, \alpha, \beta) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \left(\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)} \right).$$

Equivalently, it can be written as the form:

$$f(x; r, \alpha, \beta) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \left(\frac{\alpha(r+j)}{\alpha\beta+(r+j)} B\left(\alpha, \frac{r+j}{\beta}\right) \right) \quad (9)$$

where, $B(\cdot)$ refers to the beta function defined by:

$$B(r, s) = \frac{\Gamma(r)\Gamma(s)}{\Gamma(r+s)}, \; r, \; s > 0.$$

Next, we display the negative binomial-exponential (NB-E) distribution is a special cases of the NB-GE distribution, and there graphs of the probability mass function of a NB-GE random variable of some values of parameters in Figure 1.

Corollary 2.3 *If* $\alpha = 1$ *then the NB-GE distribution reduces to the negative binomial-exponential (NB-E) distribution with pmf given by:*

$$f(x; r, \beta) = \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \left(1 - \frac{(-(r+j))}{\beta} \right)^{-1}, \quad (10)$$

where $x = 0, 1, 2, \ldots, r$ *and* $\beta > 0$.

Proof. If $X|\lambda \sim \text{NB}(r, p = \exp -\lambda)$ and $\lambda \sim \text{GE}(\alpha = 1, \beta)$, then the pmf of X is:

$$\begin{aligned} f(x; r, \beta) &= \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(2+\frac{r+j}{\beta}\right)} \\ &= \binom{r+x-1}{x} \sum_{j=0}^{x} \binom{x}{j} (-1)^j \left(1 - \frac{(-(r+j))}{\beta} \right)^{-1}. \end{aligned}$$

From Corollary 2.3, we find the negative binomial-exponential distribution displayed in (10), which introduced by Panger and Willmot [9].

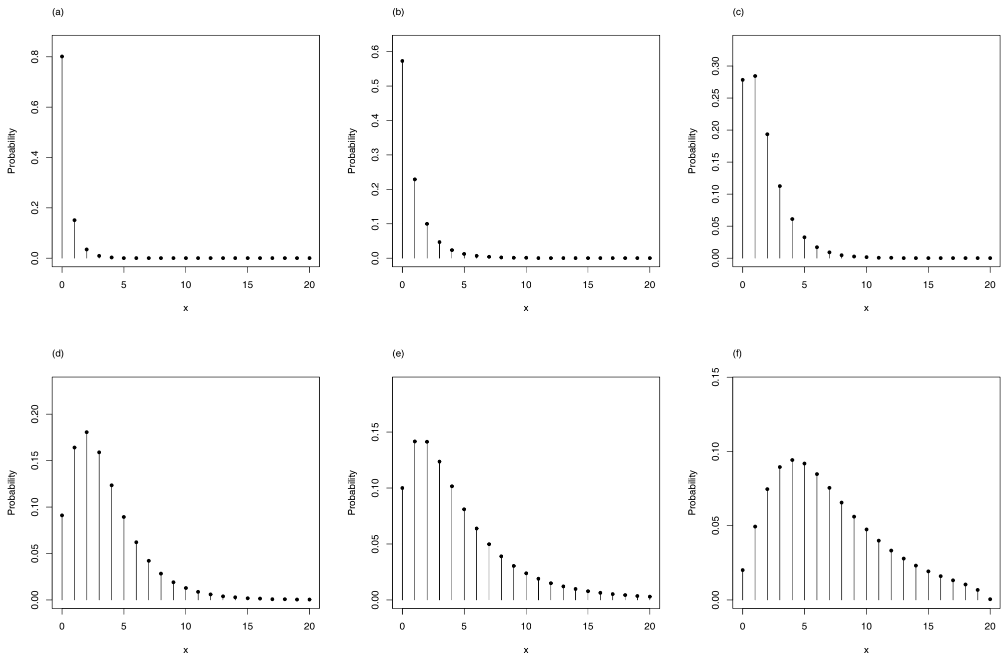

Figure 1: The probability mass function of a NB-GE random variable (X) of some values of parameters: (a) r=1, α=5, β=10, (b) r=1, α=10, β=5, (c) r=10, α=5, β=15, (d) r=10, α=10, β=10, (e) r=10, α=3, β=5 and (f) r=15, α=6, β=6.

2.2 Properties of the NB-GE distribution

The first result of this section gives the factorial moment of the NB-GE distribution. Including, this new mixed distribution have the NB-E distribution which special cases. We hardly need to emphasize the necessity and importance of factorial moment in any statistical analysis especially in applied work. Some of the most importance features and characteristics of a distribution can be studied through factorial moments (eg., mean, variance, skewness, and kurtosis).

Theorem 2.4 *If X ~NB-GE(r,α,β), then the factorial moment of order k of X is given by:*

$$\mu_{[k]}(X) = \frac{\Gamma(r+k)}{\Gamma(r)} \sum_{j=0}^{k} \binom{k}{j} (-1)^j \left(\frac{\Gamma(\alpha+1)\Gamma\left(1-\frac{k-j}{\beta}\right)}{\Gamma\left(\alpha-\frac{k-j}{\beta}+1\right)} \right), \tag{11}$$

where $x = 0, 1, 2, \ldots$ for r, α and $\beta > 0$.

Proof. If $X|\lambda \sim \text{NB}(r, p = \exp(-\lambda))$ in (1) and $\lambda \sim \text{GE}(\alpha, \beta)$, then the factorial moment of order k of X can be obtained by:

$$\mu_{[k]}(X) = E_\lambda[\mu_k(X|\lambda)].$$

The factorial moment of order k of a NB distribution in (2), $\mu_{[k]}(X)$ becomes:

$$\mu_{[k]}(X) = E_\lambda\left(\frac{\Gamma(r+k)}{\Gamma(r)}\frac{(1-e^{-\lambda})^k}{e^{-\lambda k}}\right) = \frac{\Gamma(r+k)}{\Gamma(r)}E_\lambda(e^\lambda - 1)^k.$$

Using a binomial expansion of $(e^\lambda - 1)^k$, then shows that $\mu_{[k]}(X)$ can be written as:

$$\begin{aligned}\mu_{[k]}(X) &= \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j E_\lambda(e^{\lambda(k-j)}) \\ &= \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j M_\lambda(k-j).\end{aligned}$$

From the moment generating function of GE distribution (4) with $t = k - j$, we have finally that $\mu_{[k]}(X)$ which can be written as:

$$\mu_{[k]}(X) = \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\left(\frac{\Gamma(\alpha+1)\Gamma\left(1-\frac{k-j}{\beta}\right)}{\Gamma\left(\alpha-\frac{k-j}{\beta}+1\right)}\right).$$

Corollary 2.5 *If $\alpha = 1$ then the factorial moment of order k of negative binomial-generalize exponential reduces to:*

$$\mu_{[k]}(X) = \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\left(1-\frac{k-j}{\beta}\right)^{-1},$$

where $x = 0, 1, 2, \ldots\ , r$ and $\beta > 0$.

Proof. Substituting $\alpha = 1$ into (11), we get:

$$\begin{aligned}\mu_{[k]}(X) &= \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\frac{\Gamma(2)\Gamma\left(1-\frac{k-j}{\beta}\right)}{\Gamma\left(2-\frac{k-j}{\beta}\right)} \\ &= \binom{r+x-1}{x}\sum_{j=0}^{x}\binom{x}{j}(-1)^j\frac{\left(\frac{r+j}{\beta}\right)!}{\left(1+\frac{r+j}{\beta}\right)!} \\ &= \frac{\Gamma(r+k)}{\Gamma(r)}\sum_{j=0}^{k}\binom{k}{j}(-1)^j\left(1-\frac{k-j}{\beta}\right)^{-1}.\end{aligned}$$

From the factorial moments of NB-GE distribution, it is straightforward to deduce the first four moments given in (12) - (15), variance (16), skewness (17) and kurtosis (18);

$$\begin{aligned}
E(X) &= r\left(\delta_1 - 1\right) && (12)\\
E(X^2) &= (r^2+r)\delta_2 - (2r^2+r)\delta_1 + r^2 && (13)\\
E(X^3) &= (r^3+3r^2+2r)\delta_3 - (3r^3+6r^2+3r)\delta_2 \\
&\quad +(3r^3+3r^2+r)\delta_1 - r^3 && (14)\\
E(X^4) &= (r^4+6r^3+11r^2+6r)\delta_4 - (4r^4+18r^3+26r^2+12r)\delta_3 + r^4 \\
&\quad +(6r^4+18r^3+19r^2+7r)\delta_2 - (4r^4+6r^3+4r^2+r)\delta_1 && (15)\\
Var(X) &= E(X^2) - \left(E(X)\right)^2 \\
&= (r^2+r)\delta_2 - r\delta_1\left(1+r\delta_1\right) && (16)\\
\text{Skewness}(X) &= \left[E(X^3) - 3E(X^2)E(X) + 3E(X)\left(E(X)\right)^2 - \left(E(X)\right)^3\right]/\sigma^3 \\
&= \left[(r^3+3r^2+2r)\delta_3 - (3r^2+3r)\delta_2 + r\delta_1 + 3r^2\left(\delta_1\right)^2\right. \\
&\quad \left. -(3r^3+3r^2)\delta_1\delta_2 + 2r^3\left(\delta_1\right)^3\right]/\sigma^3 && (17)\\
\text{Kurtosis}(X) &= \left[E(X^4) - 4E(X^3)E(X) + 6E(X^2)\left(E(X)\right)^2\right. \\
&\quad \left. -4E(X)\left(E(X)\right)^3 + \left(E(X)\right)^4\right]/\sigma^4 \\
&= \left[(r^4+6r^3+11r^2+6r)\delta_4 - (6r^3+18r^2+12r)\delta_3 - 3r^4\left(\delta_1\right)^4\right. \\
&\quad +(7r^2+7r)\delta_2 - 4r\delta_1 - 6r^3\left(\delta_1\right)^3 + (12r^3+12r^2)\delta_1\delta_2 \\
&\quad \left. -(4r^4+12r^3+8r^2)\delta_1\delta_3 + (6r^4+6r^3)\left(\delta_1\right)^2\delta_2\right]/\sigma^4 && (18)
\end{aligned}$$

where, $\delta_1 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{1}{\beta})}{\Gamma(\alpha-\frac{1}{\beta}+1)}$, $\delta_2 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{2}{\beta})}{\Gamma(\alpha-\frac{2}{\beta}+1)}$, $\delta_3 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{3}{\beta})}{\Gamma(\alpha-\frac{3}{\beta}+1)}$, $\delta_4 = \frac{\Gamma(\alpha+1)\Gamma(1-\frac{4}{\beta})}{\Gamma(\alpha-\frac{4}{\beta}+1)}$ and $\sigma = \sqrt{\mathrm{Var}(X)}$.

2.3 Parameters estimation

The estimation of parameters for NB-GE distribution via the MLE method procedure will be discussed. The likelihood function of the NB-GE(r, α, β) is given by:

$$L(r,\alpha,\beta) = \prod_{i=1}^{n} \binom{r+x_i-1}{x_i} \sum_{j=0}^{x_i} \binom{x_i}{j} (-1)^j \left(\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+11\right)} \right) \quad (19)$$

with corresponding log-likelihood function:

$$\begin{aligned}\mathcal{L}(r,\alpha,\beta) &= \log L(r,\alpha,\beta)\\ &= \sum_{i=1}^{n}\log\left(\Gamma(r+x_i)-\Gamma(r)-\Gamma(x_i+1)\right)+\\ &\quad \sum_{i=1}^{n}\log\left(\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}\right). \qquad (20)\end{aligned}$$

The first order conditions for finding the optimal values of the parameters obtained by differentiating (20) with respect to r, α and β give rise to the following differential equations:

$$\begin{aligned}&\frac{\partial}{\partial r}\mathcal{L}(r,\alpha,\beta)\\ &= \sum_{i=1}^{n}\psi(r+x_i)-n\psi(r)+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\Gamma(\alpha+1)\frac{\partial}{\partial r}\left(\frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}}\right)\\ &= \sum_{i=1}^{n}\psi(r+x_i)-n\psi(r)+\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\partial}{\partial r}\left(\frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}}\right) \qquad (21)\end{aligned}$$

where $\psi(x)=\frac{\Gamma'(x)}{\Gamma(x)}$ is the digamma function [1],

$$\frac{\partial}{\partial\alpha}\mathcal{L}(r,\alpha,\beta)=\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\partial}{\partial\alpha}\left(\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\Gamma(\alpha+1)\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}}\right) \qquad (22)$$

and

$$\frac{\partial}{\partial\beta}\mathcal{L}(r,\alpha,\beta)=\sum_{i=1}^{n}\left(\frac{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\partial}{\partial\beta}\left(\frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}\right)}{\sum_{j=0}^{x_i}\binom{x_i}{j}(-1)^j\frac{\Gamma\left(1+\frac{r+j}{\beta}\right)}{\Gamma\left(\alpha+\frac{r+j}{\beta}+1\right)}}\right). \qquad (23)$$

In solving equation with algebraic for such as equation in (21)-(23), which have difficult and complicated. In this paper we use Newton-Raphson method.

The Newton-Raphson method is a powerful technique for solving equations numerically. Like so much of the differential calculus, it is based on the simple idea of linear approximation [2]. Thus, we estimated parameters of distribution with estimating (21)-(23) to zero, the MLE solutions of $\hat{r}, \hat{\alpha}$ and $\hat{\beta}$ can be obtained by solving the resulting equations simultaneously using a numerical procedure with the Newton-Raphson method. Furthermore, we provide the R code for the MLE of the NB-GE distribution as follows.

2.3.1 The R code for the MLE of the NB-GE distribution

```
mlogl<-function(theta,x){
fnbge<-function(theta,x){
mm<-length(x)
k<-numeric(mm)
nbge<-function(theta,x){
if(x==0){
p<-(-log(factorial(theta[1]+x-1))+log(factorial(theta[1]-1))
    -log(gamma(theta[2]+1))-log(gamma(1+(theta[1]/theta[3])))
    +log(gamma(theta[2]+(theta[1]/theta[3])+1)))}
else if(x>0){
pp1<-(gamma(theta[2]+1)*(gamma(1+theta[1]/theta[3])))
    /(gamma(theta[2]+theta[1]/theta[3]+1))
for(j in 1:x){
p1<-((factorial(x)/(factorial(j)*factorial(x-j)))*(-1)^j)
    *(gamma(theta[2]+1))*((gamma(1+(theta[1]+j)/theta[3]))
    /(gamma(theta[2]+(theta[1]+j)/theta[3]+1)))
pp1<-pp1+p1}
p<-(-log(factorial(theta[1]+x-1)))+log(factorial(theta[1]-1))
    +log(factorial(x))-log(pp1)}
p}for(i in 1:length(x)){k[i]<-nbge(theta,x[i])}k}
sum(fnbge(theta,x))}
theta.start<-c(1,1,1)
out<-nlm(mlogl, theta.start, x=x)
r_MLE<-out$estimate[1]
a_MLE<-out$estimate[2]
b_MLE<-out$estimate[3]
```

3 Results and discussion

We used a real data set which the number of automobile liability policies in Switzerland for private cars taken from Klugman [6]. These data appear in Table 1 (first and second columns). As it can be seen in Figure 2, these data are heavily skewed to the right and overdispersed since sample variance (1.793)

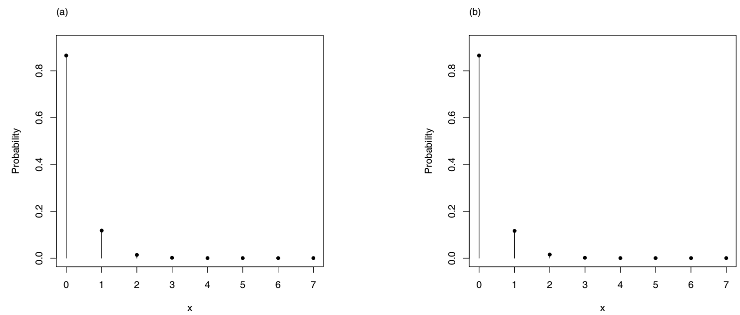

Figure 2: The probability mass function of accident data (X): (a) the empirical probability mass function of real data and (b) the probability mass function of NB-GE ($r = 2.4312, \alpha = 3.2896, \beta = 31.2786$) distribution.

Table 1: Observed and expected frequencies of accident data

Number of accidents	Observed frequency	%	Fitting distribution: Poisson	NB	NB-GE
0	103704	86.53	102629.6	103723.6	103708.8
1	14075	11.74	15922.0	13989.9	14046.8
2	1766	1.47	1235.1	1857.1	1797.8
3	255	0.21		245.2	251.7
4	45	0.04			36.0
5	6	0.01	66.3	37.2	
6	2	0.00			12.0
7+	0	0.00			
Total	11983	100	11983	11983	11983
Estimates parameters			$\hat{\lambda}= 0.1551$	$\hat{r} = 1.0327$	$\hat{r} = 2.4312$
			$\hat{p} = 0.1502$		$\hat{\alpha}= 3.2896$
					$\hat{\beta}= 31.2786$
Chi-squares			1332.30	12.12	4.26
Degrees of freedom			2	2	2
p-value			< 0.0001	0.0023	0.1188

is greater than the sample mean (1.551). Therefore, these data must be fitted by some overdispersed distribution. For this reason, we strongly believe the NB-GE distribution is suitable for them. Normally these data set are fitted by Poisson distribution, NB distribution and NB-GE distribution. The maximum likelihood method provides parameters estimation. By comparing these fitting distribution in Table 1, based on the p-value of this comparison, the results have shown that the NB-GE distribution provided better fit than the Poisson and NB distributions for the count data that have a large number of zeros.

4 Conclusions

We introduced the NB-GE distribution which is obtained by mixing the NB distribution with a GE distribution. We showed that the NB-E distribution is special cases of this new mixed distribution. We have obtained the key moments of the NB-GE distribution which includes the factorial moments, mean, variance, skewness, and kurtosis. Parameters estimation are also implemented using maximum likelihood method and the usefulness of the NB-GE distribution is illustrated by real data set.

This paper has described the application of the NB-GE distribution to data sets characterized by a large number of zeros and a heavy tail. Traditional statistical methods that have been proposed for analyzing such data sets have been found to suffer from important numerical and methodological problems. From the result that the NB-GE distribution provides a best fit for these data. In conclusion, it is believed that the NB-GE distribution may offer a very useful tool for analyzing count data characterized with a large number of zeros.

ACKNOWLEDGEMENTS. The authors would like to thank the Research Professional Development Project under the Science Achievement Scholarship of Thailand (SAST).

References

[1] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables*, New York, 1972.

[2] M.D Bressoud, *A Radical Approach to Real Analysis*, 2nd edn, Cambridge Univ Press, Cambridge, 2005.

[3] E. Gomez-D́eniz, J.M. Sarabia and E.Calderń-Ojeda, Univariate and multivariate versions of the negative binomial-inverse Gaussian distributions with applications, *Insurance Mathematics and Economics* **42** (2008), 39–49.

[4] R.D. Gupta and D. Kundu, Generalized exponential distributions, *Austral. New Zealand J. Statist* **41** (1999), 173–188.

[5] N.L Johnson and S. Kotz, *Univariate Discrete Distributions*, 2nd edn, Wiley, NewYork, 2005.

[6] S. Klugman, *Loss Models: From Data to Decisions*, 2nd edn, Wiley–Inter Science, 2004.

[7] D. Lord and R.S. Geedipally, The negative binomial–Lindley distribution as a tool for analyzing crash data characterized by a large amount of zeros, *J. Accident Analysis and Prevention*, **43** (2011), 1738–1742.

[8] S. Meng, Y. Wei and G.A. Whitmore, Accounting for individual overdispersion in a Bonus-Malus system, *Astin Bulletin*, **2**(1999), 327–337.

[9] H.H. Panjer and G.E. Willmot, Finite sum eqvluation of the negative binomail–exponential model, *Astin Bulletin* **12** (1981), 133–137.

[10] C. Pudprommarat, W. Bodhisuwan and P. Zeephongsekul, A new mixed negative binomial distribution, *J. Applied Sciences* **17** (2012), 1853–1858.

[11] W. Rainer, *Econometric Analysis of Count Data*, 3rd edn, Springer-Verlag, Berlin, Germany, 2000.

[12] H. Ranger and E. Willmot, Finite sum evaluation of the negative binomial-exponential model, *Astin Bulletin* **12** (1981), 133–137.

[13] L. Simon, Fitting negative binomial distribution by the method of maximum likelihood, *J. Casualty Actuarial Society* **17** (1961), 45–53.

[14] Z. Wang, One mixed negative binomial distribution with application, *Statistical Planning and Inference*, **141** (2011), 1153–1160.

[15] R. Winkelmann, *Econometric Analysis of Count Data*, 3rd edn, Springer-Verlag, Berlin, Germany, 2003.

[16] H. Zamani and N. Ismail, Negative binomial–Lindley distribution and its application, *J. Mathematics and Statistics* **1** (2010), 4–9.

Received: November, 2012