Quasi-negative binomial distribution: Properties and applications

Shubiao Li [a], Fang Yang [b,1], Felix Famoye [c,*], Carl Lee [c], Dennis Black [a]

[a] *Comerica Loan Center, Comerica Bank, Auburn Hills, MI, USA*
[b] *ACAS, Auto Club Group, Dearborn, MI, USA*
[c] *Department of Mathematics, Central Michigan University, Mt. Pleasant, MI, USA*

ARTICLE INFO

ABSTRACT

In this paper, a quasi-negative binomial distribution (QNBD) derived from the class of generalized Lagrangian probability distributions is studied. The negative binomial distribution is a special case of QNBD. Some properties of QNBD, including the upper tail behavior and limiting distributions, are investigated. It is shown that the moments do not exist in some situations and the limiting distribution of QNBD is the generalized Poisson distribution under certain conditions. A zero-inflated QNBD is also defined. Applications of QNBD and zero-inflated QNBD in various fields are presented and compared with some other existing distributions including Poisson, generalized Poisson and negative binomial distributions as well as their zero-inflated versions. In general, the QNBD or its zero-inflated version performs better than the other models based on the chi-square statistic and the Akaike Information Criterion, especially for the cases where the data are highly skewed, have heavy tails or excessive numbers of zeros.

1. Introduction

Count data are often encountered in real world applications. There is a very large collection of literature on how to analyze and model count data. Many discrete distributions have been developed for modeling count data. Among the commonly used distributions are the Poisson distribution, the generalized Poisson distribution (GPD) and the negative binomial distribution (NBD). For a review of the literature, one may refer to Johnson et al. (2005), Consul (1989) and Consul and Famoye (2006). The Poisson distribution has the property of equal mean and variance (equi-dispersion). However, count data often exhibit under-dispersion or over-dispersion. Over-dispersion relative to the Poisson distribution is when the sample variance is substantially in excess of the sample mean. Under-dispersion relative to the Poisson is when the sample mean is substantially in excess of the sample variance. One way to describe the dispersion characteristic is to introduce an additional parameter into the Poisson distribution for handling over- or under-dispersion phenomena.

The classical modeling on accident proneness theory and actuarial risk theory by Greenwood and Yule (1920) introduced a parameter λ for characterizing the accident intensity of an individual. By modeling the number of accidents per individual and assuming that it follows a Poisson distribution with parameter λ conditioned on a realization of a gamma random variable, Greenwood and Yule developed the distribution of the Poisson distribution compounded with the gamma distribution and found that it is the NBD.

Starting in the early 1970s, Consul and Jain (1973a,b) developed the GPD that can model both under- and over-dispersed count data. Consul and Shoukri (1985) and Consul and Famoye (1986, 1988, 1989) reported a series of studies about the

---

* Corresponding author. Tel.: +1 989 774 5497; fax: +1 989 774 2414.
 E-mail address: felix.famoye@cmich.edu (F. Famoye).
[1] New address: JP Morgan Chase, Iselin, NJ, USA.

properties, estimation and applications of GPD. Consul and Shenton (1973) showed that the GPD could be derived from the Lagrangian probability distributions of the first kind. For details on GPD, one may refer to Consul (1989) and Consul and Famoye (2006).

Li et al. (2006) defined the generalized Lagrangian distribution class and developed a class of 'quasi' discrete distributions. Li et al. (2008) defined certain mixture distributions based on Lagrangian probability models and derived some new distributions including the quasi-negative binomial distribution (QNBD). For details of quasi discrete and mixture distributions, one may refer to Li (2007). In Section 2, we present some properties of the QNBD, including the upper tail behavior of the distribution and the limiting distribution. Zero-inflated QNBD is defined in Section 3; the application of QNBD and zero-inflated QNBD is presented in Section 4.

2. Some properties of quasi-negative binomial distribution

Li et al. (2006) defined the class of generalized Lagrangian distributions, in which an extra parameter (the Lagrangian expansion point) was brought into its probability mass function. Let $f(z)$ and $g(z)$ be analytic functions, $\{D^{x-1}[(g(z))^x f'(z)]\}_{z=0} \geq 0$ and $g(0) > 0$, where $D = \partial/\partial z$. If there is a point $t > 0$, such that $f(t) > 0$ and $g(t) > 0$, for $x \in N$, where N is the set of natural numbers, then the generalized Lagrangian probability distribution of the first kind is defined as

$$P(X = x) = \begin{cases} f(0)/f(t), & x = 0 \\ \dfrac{(t/g(t))^x}{x!f(t)}\{D^{x-1}[(g(z))^x f'(z)]\}_{z=0}, & x = 1, 2, 3, \ldots. \end{cases} \tag{1}$$

Many discrete probability distributions can be derived by using specific $f(z)$ and $g(z)$ functions. For example, if $g(z) = e^{\lambda z}$ and $f(z) = e^{\theta z}$, then the probability mass function (pmf) of the generalized Lagrangian distribution is

$$P(X(t) = x) = \theta t(\theta t + \lambda tx)^{x-1}e^{-\theta t - \lambda tx}/x!, \quad \text{for } x = 0, 1, 2, \ldots, \tag{2}$$

which is the pmf of the GPD when $t = 1$. In the pmf in (2), we have $0 < \lambda < 1, \theta > 0$ and $t > 0$.

The variable t in (2) is positive and can be considered as a random variable. Li et al. (2008) derived several new distributions by assuming different distributions for t. One of the new distributions derived is the QNBD, which assumes that t follows the gamma distribution with density

$$s(t) = \beta^\alpha t^{\alpha-1}e^{-\beta t}/\Gamma(\alpha), \quad t > 0, \ \alpha > 0 \text{ and } \beta > 0.$$

A discrete random variable X is said to have a QNBD if its pmf (Li et al., 2008) is given by

$$P(X = x) = \begin{cases} \dfrac{\Gamma(x+\alpha)}{x!\Gamma(\alpha)} \dfrac{1}{1+cx}\left(\dfrac{1+cx}{1+b+cx}\right)^x\left(\dfrac{b}{1+b+cx}\right)^\alpha, & x = 0, 1, 2, \ldots \\ 0, & \text{for } x > m \text{ if } c < 0, \end{cases} \tag{3}$$

where $\alpha > 0, b > 0, m \geq [-1/c]$, and m is the largest positive integer for which $1 + mc \geq 0$. For the QNBD, the probability of zero class is $P(X = 0) = [b/(1+b)]^\alpha$ and the probability of at least one event occurring is $1 - [b/(1+b)]^\alpha$. The QNBD in (3) reduces to the NBD when the parameter $c = 0$. This additional parameter c of QNBD can be used to describe the pattern of non-zero count. When $c = 0$ and α is an integer, $b/(1+b)$ is the NBD probability of "success" and x is the number of failures to achieve α successes. For the QNBD in (3), the probability of "success" is $b/(1+b+cx)$, which depends on the value of x and the parameters b and c.

2.1. The upper tail property of QNBD

The first property of QNBD investigated is the upper tail behavior of the distribution. Two functions $h(x)$ and $\omega(x)$ are said to have the same order, denoted by $h(x) \sim \omega(x)$, if the limit of $h(x)/\omega(x)$ as x approaches infinity is a non-zero constant k. That is, $\lim_{x\to\infty} h(x)/\omega(x) = k \neq 0$, where k is a constant (Walter, 1976). If the function $h(x)$ is a pmf of a random variable, then the upper tail behavior of $h(x)$ can be studied through $\omega(x)$. The following theorem shows that the upper tail of QNBD when $c > 0$ has the same order as the function $\omega(x) = x^{-2}$.

Theorem 1. Let $h(x)$ be the pmf of QNBD in (3) and let $\omega(x) = x^{-2}$, then $h(x)$ and $\omega(x)$ have the same asymptotic order when $c > 0$.

Proof. By using the asymptotic Stirling's formula, $\Gamma(x + 1) \sim \sqrt{2\pi}x^{x+1/2}e^{-x}$ (Abramowitz and Stegun, 1972, page 257), we have

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{\sqrt{2\pi}x^{x+1/2}e^{-x}} = 1 \quad \text{and} \quad \lim_{x\to\infty} \frac{\Gamma(x+\alpha)}{\sqrt{2\pi}(x+\alpha-1)^{x+\alpha-1/2}e^{-(x+\alpha-1)}} = 1.$$