Journal of Statistical Planning and Inference 141 (2011) 1153–1160

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

One mixed negative binomial distribution with application

Zhaoliang Wang

School of Mathematics and Information Science, Henan Polytechnic University, Jiaozuo 454000, China

ARTICLE INFO

Article history:
Received 9 December 2009
Received in revised form
9 September 2010
Accepted 16 September 2010
Available online 24 September 2010

Keywords:
Over-dispersion
Mixture
Goodness-of-fit
Poisson
Negative binomial

ABSTRACT

In this paper we do some research on a three-parameter distribution which is called beta-negative binomial (BNB) distribution, a beta mixture of negative binomial (NB) distribution. The closed form and the factorial moment of the BNB distribution are derived. In addition, we present the recursion on the pdf of BNB stopped-sum distribution, and make stochastic comparison between BNB and NB distributions. Furthermore, we have shown that BNB distribution has heavier tail than NB distribution. The application of BNB distribution is carried out on one sample of insurance data. Based on the results, we have shown that the BNB provides a better fit compared to the Poisson and the NB for count data.

1. Introduction

Poisson distribution is usually employed to fit the count data in practice, but theoretical prediction may not match empirical observations for moments of higher order due to the only free parameter, which does not allow the variance to be adjusted independently of the mean. Furthermore, one important feature of Poisson family is the one unit variance-to-mean ratio called dispersion index which is used to measure the departure from Poisson distribution. The observed count data often displays features like over-dispersion, and sometimes even heavy tail for populations under study are frequently heterogeneous, which is common in applied data analysis. Statistical procedures must account for these sources of variability in modelling the variation in observed counts.

The Poisson distribution plays an important role in the modelling process even if it is not applicable, because it is often used to derive more realistic models that meet characteristics of observed data. Over-dispersion was noted by Greenwood and Yule (1920), which suggested a model in which the mean of Poisson distribution has a gamma distribution. Actually, this leads to the negative binomial (NB) distribution. Some mixed Poisson distributions with additional free parameters are used in practice (see Karlis and Xekalaki, 2005).

NB distribution has become increasingly popular as a more flexible alternative to Poisson distribution, especially when it is doubtful whether the strict requirements (particularly, the independence) for Poisson distribution could be satisfied. For various extensions, modifications as well as some applications on Poisson and NB distribution, please refer to Johnson et al. (2005). But NB distribution is better for over-dispersed count data that are not necessarily heavy-tailed. The extremely heavy tail implies over-dispersion, but the converse does not hold. Let μ and σ^2 be the mean and the variance. Thus, we can measure the tail heaviness relative to NB by the tail index $\Delta=\gamma-\gamma_{NB}$, where $\gamma_{NB}=(2\sigma^2-\mu)/(\mu\sigma)$ is the skewness of NB. Positive or negative value of the index indicates longer or shorter tail relative to NB. Assume sample mean, sample variance and sample skewness are $\hat{\mu},\hat{\sigma}^2$ and $\hat{\gamma}$, respectively, then γ_{NB} can be estimated as $\hat{\gamma}_{NB}=(2\hat{\sigma}^2-\hat{\mu})/(\hat{\mu}\hat{\sigma})$, and the sample tail index is $\hat{\Delta}=\hat{\gamma}-\hat{\gamma}_{NB}$, which provides information regarding whether a distribution with short or long tail should be used.

E-mail address: mathwangzhl@yahoo.cn

0378-3758/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.jspi.2010.09.020

This paper considers the mixture of NB distribution instead of NB itself and Poisson distribution: the probability of success is assumed to have a beta distribution. The incurred distribution is called beta-negative binomial (BNB) distribution, which introduces one more free parameter and is more flexible in fitting count data. The rest of this paper is organized as follows: Section 2 presents some basic characteristics of BNB distribution. In Section 3, the BNB stopped-sum distribution is presented. We obtain a recurrence for the probability mass function (pmf) for BNB distribution and then an integral equation is derived for probability density function (pdf) for BNB stopped-sum distribution. In Section 4, we present stochastic comparison between BNB and NB distributions, and show that BNB distribution has heavier tail than NB. Finally, one sample of insurance data is provided in Section 5 to illustrate the greater versatility of BNB distribution.

2. Basic characteristics

A random variable Y of NB distribution (denoted as $Y\sim \mathrm{NB}(r,p)$) has its pmf

$$\Pr(Y=k)=\frac{\Gamma(k+r)}{k!\Gamma(r)}p^r(1-p)^k,\quad k=0,1,2,\ldots, \tag{2.1}$$

where $0<p<1$, $r>0$ is not necessarily an integer and Γ denotes the gamma function. A random variable P with beta distribution (denoted as $P\sim \mathrm{Beta}(\alpha,\beta)$) has its pdf

$$f(p)=\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)},\quad 0<p<1, \tag{2.2}$$

where $\alpha,\beta>0$ and B denotes the beta function.

A random variable Z has BNB distribution if it satisfies the stochastic representation: $Z|p\sim \mathrm{NB}(r,p)$ and $p\sim \mathrm{Beta}(\alpha,\beta)$. Throughout this study, we will use the notation $Z\sim \mathrm{BNB}(r,\alpha,\beta)$ as a reference for the BNB distribution. Since NB distribution is a mixture of Poisson distribution, BNB distribution can also be regarded as a kind of mixed Poisson distribution.

2.1. Probability mass and moments

In this part, the closed form and the factorial moment of the BNB distribution are given. If $Z|p\sim \mathrm{NB}(r,p)$ and $p\sim \mathrm{Beta}(\alpha,\beta)$, according to (2.1) and (2.2), then the pmf of Z is

$$\Pr(Z=k)=p_k=\int_0^1 \Pr(Z=k|p)f(p)\,\mathrm{d}p=\int_0^1\frac{\Gamma(k+r)}{k!\Gamma(r)}p^r(1-p)^k\frac{p^{\alpha-1}(1-p)^{\beta-1}}{B(\alpha,\beta)}\,\mathrm{d}p=\frac{\Gamma(k+r)}{k!\Gamma(r)}\frac{B(\alpha+r,\beta+k)}{B(\alpha,\beta)},\quad k=0,1,2,\ldots. \tag{2.3}$$

Hence the moment generating function (mgf) $g(t)=\mathrm{E}(t^Z)$ can be obtained as

$$g(t)=H(r,\beta;r+\alpha+\beta,t)\Pr(Z=0), \tag{2.4}$$

where $H(a,b;c,t)=\sum_{n=0}^{\infty}((a)_n(b)_n/(c)_n)t^n/n!$ is the hypergeometric function (see Abramowitz and Stegun, 1964 (p.558)), and $(x)_n=\Gamma(x+n)/\Gamma(x)$ for all $x>0$.

By taking the derivative on $g(t)$ with respect to t in (2.4), we have

$$g^{(k)}(t)=\frac{(r)_k(\beta)_k}{(r+\alpha+\beta)_k}H(r+k,\beta+k;r+\alpha+\beta+k,t)\Pr(Z=0),\quad \alpha-k>0.$$

Since the factorial moment $m_{[k]}(Z)=\mathrm{E}[\Pi_{j=0}^{k-1}(Z-j)]=g^{(k)}(1)$, for $k\geq 1$, it follows naturally from $H(a,b;c,1)=\Gamma(c)\Gamma(c-a-b)/\Gamma(c-a)\Gamma(c-b)$, for $c-a-b>0$ that

$$m_{[k]}(Z)=\frac{\Gamma(r+k)}{\Gamma(r)}\frac{B(\alpha-k,\beta+k)}{B(\alpha,\beta)},\quad k<\alpha. \tag{2.5}$$

The moments at the origin can be easily obtained by $m_n=\sum_{k=1}^{n}S_n^k m_{[k]}$, where S_n^k stands for the Stirling numbers of the second kind (see Rennie and Dobson, 1969).

As r tends to zero ($r\to 0^+$), the BNB converges to a distribution degenerated at zero (i.e. $\Pr(Z=0)=1$). Furthermore, by using Stirling's formula, $\Gamma(s+1)=\sqrt{2\pi s}(s/\mathrm{e})^s\mathrm{e}^{\theta/12s}$ for $s>0$ and $\theta\in(0,1)$, we get $\lim_{r\to+\infty}p_k=0$ for any k. Clearly, the parameter r and β are exchangeable. Thus, r and β can be interpreted as degeneration parameters.

Consider a sequence of BNB distributions where parameter α tends to infinity in such a way as to keep the mean of the beta distribution constant. Denoting this mean $\alpha/(\alpha+\beta)=\omega$, the parameter β will have to be $(1-\omega)/\omega\alpha$. Under this parametrization, the BNB pmf in (2.3) will become

$$p_k=\binom{k+r-1}{k}\frac{\Gamma(\alpha+r)\Gamma\left(\frac{1-\omega}{\omega}\alpha+k\right)\Gamma\left(\frac{\alpha}{\omega}\right)}{\Gamma\left(\frac{\alpha}{\omega}+r+k\right)\Gamma(\alpha)\Gamma\left(\frac{1-\omega}{\omega}\alpha\right)},\quad k=0,1,2,\ldots.$$