$\mathcal{A}\mathcal{P}$
ijpam.eu

NEGATIVE BINOMIAL APPROXIMATION TO
THE BETA BINOMIAL DISTRIBUTION

K. Teerapabolarn

Department of Mathematics
Faculty of Science
Burapha University
Chonburi, 20131, THAILAND

Abstract: This paper determines a bound on the approximation of the beta binomial distribution with parameters n, α and β by a negative binomial distribution with parameters α and $\frac{\alpha+\beta}{\alpha+\beta+n}$. With this bound, it is indicated that the beta binomial distribution can be well approximated by the negative binomial distribution when β is large.

AMS Subject Classification: 62E17, 60F05
Key Words: beta binomial distribution, negative binomial approximation, Stein's method, w-function associated with the beta binomial random variable

1. Introduction

The beta binomial random variable X with parameters $n \in \mathbb{N}$, $\alpha > 0$, and $\beta > 0$, with distribution written $\mathcal{BB}_{n,\alpha,\beta}$, has probabilities

$$p_X(x) = \binom{n}{x} \frac{B(x+\alpha, n-x+\beta)}{B(\alpha, \beta)}, \ x = 0, 1, ..., n, \tag{1.1}$$

and has mean $\mu = \frac{n\alpha}{\alpha+\beta}$ and variance $\sigma^2 = \frac{n\alpha\beta(n+\alpha+\beta)}{(\alpha+\beta)^2(1+\alpha+\beta)}$, where B is the complete beta function. It is well-known that the beta binomial can be approximated by some discrete distributions, such as Poisson and binomial distributions, provided that their parameters are given under appropriate conditions.

Received: August 18, 2014

For example, Teerapabolarn [3] gave a bound for the total variation distance between the binomial and beta binomial distributions as follows:

$$d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{B}_{n,p}) \leq \left(1 - p^{n+1} - q^{n+1}\right) \frac{n(n-1)}{(n+1)(1+\alpha+\beta)} \qquad (1.2)$$

for $A \subseteq \{0, ..., n\}$, where $d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{B}_{n,p}) = \sup_A |\mathcal{BB}_{n,\alpha,\beta}\{A\} - \mathcal{B}_{n,p}\{A\}|$ and $\mathcal{B}_{n,p}$ is the binomial distribution with parameters n and $p = 1 - q = \frac{\alpha}{\alpha+\beta}$. Later, he also gave a bound for the total variation distance between the Poisson and beta binomial distributions [4] as follows:

$$d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{P}_\lambda) \leq \left(1 - e^{-\lambda}\right) \frac{(\alpha+1)(\alpha+\beta) - n\beta}{(\alpha+\beta)(\alpha+\beta+1)} \qquad (1.3)$$

for $A \subseteq \mathbb{N} \cup \{0\}$, where $\mathcal{P}_\lambda$ is the Poisson distribution with mean $\lambda = \frac{n\alpha}{\alpha+\beta}$ and $\alpha \geq (n-1)$. Let us consider (1.1), we have

$$p_X(x) = \frac{\Gamma(\alpha+x)n!\Gamma(\beta+n-x)\Gamma(\alpha+\beta)}{x!\Gamma(\alpha)(n-x)!\Gamma(\alpha+\beta+n)\Gamma(\beta)}. \qquad (1.4)$$

It is observed that $\alpha + \beta \to \infty$ while $\frac{n}{\alpha+\beta}$ remains constant, then the beta binomial distribution with parameters n, α and β converges to a negative binomial distribution with parameters α and $\frac{\alpha+\beta}{\alpha+\beta+n}$. Therefore, it is reasonable to approximate the beta binomial distribution by the negative binomial with these parameters when $\alpha + \beta$ is sufficiently large. In this paper, we use Stein's method and the w-function associated with the random variable X to give a bound for the total variation distance between the beta binomial and negative binomial distributions.

2. Method

The following lemma is the w-function associated with the beta binomial random variable X, which is directly obtained from [3].

Lemma 2.1. *We have*

$$w(x) = \frac{(n-x)(\alpha+x)}{(\alpha+\beta)\sigma^2}, \quad x = 0, 1, ..., n, \qquad (2.1)$$

where $\sigma^2 = \frac{n\alpha\beta(n+\alpha+\beta)}{(\alpha+\beta)^2(1+\alpha+\beta)}$.

The next relation is an important property to obtain the desired result, which was stated by [2]. If a non-negative integer-valued random variable X has $p_X(x) > 0$ for every x in support of X and $0 < \sigma^2 = \mathbb{V}ar(X) < \infty$, then

$$\mathbb{C}ov(X, g(X)) = \sigma^2 \mathbb{E}[w(X)\Delta g(X)], \tag{2.2}$$

for any function $g : \mathbb{N} \cup \{0\} \to \mathbb{R}$ for which $\mathbb{E}|w(X)\Delta g(X)| < \infty$, where $\Delta g(x) = g(x+1) - g(x)$ and $\mathbb{E}[w(X)] = 1$.

For Stein's method in the negative binomial approximation, following [1], it is applied for real number $r > 0$ and $0 < p = 1 - q < 1$, for every subset A of $\mathbb{N} \cup \{0\}$ and the bounded real valued function $g = g_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ defined in [1], yields

$$\mathcal{BB}_{n,\alpha,\beta}\{A\} - \mathcal{NB}_{r,p}\{A\} = \mathbb{E}[q(r + X)g(X + 1) - Xg(X)]. \tag{2.3}$$

For any subset A of $\mathbb{N} \cup \{0\}$, [1] showed that

$$\sup_A |\Delta g(x)| \le \frac{1 - p^r}{rq}. \tag{2.4}$$

3. Result

The following theorem gives a result in the negative binomial approximation to the beta binomial distribution.

Theorem 3.1. *For $r = \alpha$, $p = 1 - q = \frac{\alpha+\beta}{\alpha+\beta+n}$ and $A \subseteq \mathbb{N} \cup \{0\}$, then*

$$d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{NB}_{\alpha,p}) \le (1 - p^\alpha)\frac{(\alpha + 1)(\alpha + \beta + n)}{(\alpha + \beta)(\alpha + \beta + 1)}. \tag{3.1}$$

Proof. From (2.3), we have

$$\begin{aligned}
\mathcal{BB}_{n,\alpha,\beta}\{A\} - \mathcal{NB}_{r,p}\{A\} &= \mathbb{E}[rqg(X + 1) + qX\Delta g(X) - pXg(X)] \\
&= rq\mathbb{E}[g(X + 1)] + q\mathbb{E}[X\Delta g(X)] - p\mathbb{E}[Xg(X)] \\
&= rq\mathbb{E}[\Delta g(X)] + q\mathbb{E}[X\Delta g(X)] - p\mathbb{C}ov(X, g(X)).
\end{aligned}$$

Using Lemma 2.1 and (2.2), we have $E|w(X)\Delta f(X)| < \infty$. From which it follows that

$$\begin{aligned}
d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{NB}_{\alpha,p}) &= \left|rq\mathbb{E}[\Delta g(X)] + q\mathbb{E}[X\Delta g(X)] - p\mathbb{E}[\sigma^2 w(X)\Delta g(X)]\right| \\
&\le \mathbb{E}\{|(r + X)q - \sigma^2 w(X)p||\Delta g(X)|\}
\end{aligned}$$

$$= \mathbb{E}\left\{\frac{(\alpha + X)X}{\alpha + \beta + n}|\Delta g(X)|\right\} \ (\geq 0).$$

Following (2.4), we obtain

$$d_{TV}(\mathcal{BB}_{n,\alpha,\beta}, \mathcal{NB}_{\alpha,p}) \leq \mathbb{E}\{[(r + X)q - \sigma^2 w(X)p]|\Delta g(X)|\}$$
$$= (\mu - \sigma^2 p)\mathbb{E}|\Delta g(X)|$$
$$\leq (1 - p^\alpha)\frac{(\alpha + 1)(\alpha + \beta + n)}{(\alpha + \beta)(\alpha + \beta + 1)},$$

which completes the proof. $\square$

If $\alpha = 1$, then $p = \frac{\beta+1}{\beta+n+1}$. Thus the approximation in Theorem 3.1 is an approximation of the beta binomial distribution with parameters n and β by a geometric distribution, $\mathcal{G}_p$, with parameter $p = \frac{\beta+1}{\beta+n+1}$.

Corollary 3.1. *For geometric approximation, we have*

$$d_{TV}(\mathcal{BB}_{n,\beta}, \mathcal{G}_{\frac{+1}{+n+1}}) \leq \frac{2n}{(\beta + 1)(\beta + 2)}. \tag{3.2}$$

4. Concluding Remarks

In the present study, a bound for the total variation distance between the beta binomial and negative binomial distributions is derived by using Stein's method and the w-function associated with the beta binomial random variable. With this bound, it is observed that the result in Theorem 3.1 gives a good approximation when β is large, that is, the negative binomial distribution with parameters α and $\frac{\alpha+\beta}{\alpha+\beta+n}$ can be used as an approximation of the beta binomial distribution with parameters n, α and β when β is sufficiently large. Additionally, it has no any conditions on their parameters of the approximation, which is similar to that mentioned in (1.2) and is different from the Poisson approximation as mentioned in (3.1).

References

[1] T.C. Brown, M.J. Phillips, Negative binomial approximation with Stein's method, *Meth. Comp. Appl. Probab.*, **1** (1999), 407–421.

[2] T. Cacoullos, V. Papathanasion, Characterization of distributions by variance bounds, *Statist. Probab. Lett.*, **7** (1989), 351–356.

[3] K. Teerapabolarn, A bound on the binomial approximation to the beta binomial distribution, Int. Math. Forum, **3**(2008), 1355–1358.

[4] K. Teerapabolarn, Poisson approximation to the beta binomial distribution, Int. Math. Forum, **4**(2009), 1253–1256.