A Bound on the Binomial Approximation to the Beta Binomial Distribution

K. Teerapabolarn

Department of Mathematics, Faculty of Science
Burapha University, Chonburi 20131, Thailand
kanint@buu.ac.th

Abstract

We use the $w$-function and the Stein identity to give a result on the binomial approximation to the beta-binomial distribution in terms of the total variation distance between the beta-binomial and binomial distributions and its upper bound.

Mathematics Subject Classification: Primary 60F05

Keywords: Beta-binomial distribution; binomial distribution; binomial approximation; Stein identity; $w$-function

1 Introduction

It is well known that the binomial distribution is the distribution of the number of successes that occur in $n$ independent trials with the probability of success in each trial is $p$. The binomial distribution with parameters $n > 0$ and $0 < p < 1$, written $B(n, p)$, has probabilities

$$p(k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, 1, ..., n.$$

The beta binomial distribution is a binomial distribution whose probability of success $p$ follows a beta distribution with shape parameters $\alpha$ and $\beta$. In other words, if the probability of success $p$ of a binomial distribution has a beta distribution with shape parameters $\alpha$ and $\beta$, then the resulting distribution is referred to as the beta binomial distribution with parameters $\alpha$, $\beta$ and $n$. For a standard binomial distribution, $p$ is usually assumed to be fixed for successive trials. For the beta binomial distribution the value of $p$ changes for each trial.
Let $X$ be the beta binomial random variable with distribution written by

$$BB(\alpha, \beta, n) = \left\{ p(k) = \binom{n}{k} \frac{B(k + \alpha, n - k + \beta)}{B(\alpha, \beta)}, \; k = 0, 1, ..., n \right\},$$

where $B$ is the complete beta function, $\alpha, \beta$ are positive real numbers and $n$ is a positive integer. Its mean and variance are $\mu = \frac{n\alpha}{\alpha + \beta}$ and $\sigma^2 = \frac{n\alpha\beta(n + \alpha + \beta)}{(\alpha + \beta)^2(1 + \alpha + \beta)}$, respectively, see Weisstein [5].

Note that the beta binomial distribution is obtained from a binomial distribution as mentioned above, thus it is natural to speculate that the beta binomial distribution is approximately the binomial distribution. In this paper, we use the $w$-function associated with the random variable $X$ together with the Stein identity to give an upper bound for the total variation distance between the beta binomial and binomial distributions.

## 2 Main result

This section we use the $w$-function associated with the beta binomial random variable $X$ and the Stein identity for binomial distribution to give the main result of the binomial approximation to the beta binomial distribution. For the $w$-function, Majserowska [4] adapted the relation of $w$-function associated with the random variable $X$ (Cacoullos and Papathanasiou [2]) to be the recurrence relation of $w$-function in the form of

$$w(k + 1) = \frac{p(k)}{p(k + 1)}w(k) - \frac{\mu - (k + 1)}{\sigma^2} \geq 0, \; k = 0, 1, ..., n,$$

where $w(0) = \frac{\mu}{\sigma^2}$ and $\mu$ and $\sigma^2$ are mean and variance of $X$.

The following proposition is an important property of the $w$-function associated with the beta binomial distribution.

**Proposition 2.1.** Let $w(X)$ be the $w$-function associated with the beta binomial random variable $X$ and $p(k) > 0$ for every $0 \leq k \leq n$. Then we have

$$w(k) = \frac{(n - k)(\alpha + k)}{(\alpha + \beta)\sigma^2}, \; k = 0, 1, ..., n,$$

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

**Proof.** Following (1.1), we have

$$\frac{p(k - 1)}{p(k)} = \frac{k(\beta + n - k)}{(n - k + 1)(\alpha + k - 1)}, \; k = 1, ..., n.$$
Using (2.1), the recurrence relation of the $w$-function can be expressed in the form

$$w(k) = \frac{n\alpha}{\sigma^2(\alpha + \beta)} + w(k - 1) \frac{k(\beta + n - k)}{(n - k + 1)(\alpha + k - 1)} - \frac{k}{\sigma^2}, \ k = 1, \ldots, n,$$

where $w(0) = \frac{n\alpha}{\sigma^2(\alpha + \beta)}$.

Therefore, we have

$$w(1) = \frac{(n - 1)(\alpha + 1)}{(\alpha + \beta)\sigma^2}, \ w(2) = \frac{(n - 2)(\alpha + 2)}{(\alpha + \beta)\sigma^2}, \ldots, w(n) = \frac{(n - n)(\alpha + n)}{(\alpha + \beta)\sigma^2},$$

which gives (2.2).

For the Stein identity, we can apply the Stein identity in Barbour et al. [1] on pp. 188-189, i.e. for fixed parameters $n \geq 1$ and $p = 1 - q \in (0, 1)$, every subset $A$ of $\{0, \ldots, n\}$ and the bounded real valued function $f = f_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ (defined as in [1]) the Stein identity for binomial case is given by

$$\mathbb{B}(\alpha, \beta, n)(A) - \mathcal{B}(n, p)(A) = E[(n - X)p f(X + 1) - q X f(X)]. \quad (2.3)$$

For any subset $A$ of $\{0, \ldots, n\}$, Ehm [3] showed that

$$\sup_{k, A} |\Delta f(k)| = \sup_{k, A} |f(k + 1) - f(k)| \leq \frac{(1 - p^{n+1} - q^{n+1})}{(n + 1)pq}. \quad (2.4)$$

The theorem below gives an upper bound on the binomial approximation to the beta binomial distribution.

**Theorem 2.1.** For $A \subseteq \mathbb{N} \cup \{0\}$, if $p = \frac{\alpha}{\alpha + \beta}$, then

$$d_{TV}(\mathbb{B}(\alpha, \beta, n), \mathcal{B}(n, p)) \leq (1 - p^{n+1} - q^{n+1}) \frac{n(n - 1)}{(n + 1)(1 + \alpha + \beta)}. \quad (2.5)$$

where $d_{TV}(\mathbb{B}(\alpha, \beta, n), \mathcal{B}(n, p)) = \sup_A |\mathbb{B}(\alpha, \beta, n)(A) - \mathcal{B}(n, p)(A)|$.

**Proof.** It can be seen that

$$E[(n - X)p f(X + 1) - q X f(X)] = E[\mu f(X + 1) - pX\Delta f(X) - X f(X)]$$

$$= E[\mu f(X + 1)] - pE[X\Delta f(X)] - E[X f(X)]$$

$$= E[\mu f(X + 1)] - pE[X\Delta f(X)] - \text{Cov}(X, f(X)) - E[\mu f(X)]$$

$$= E[\mu\Delta f(X)] - pE[X\Delta f(X)]$$

$$- \sigma^2 E[w(X)\Delta f(X)] \tag{by Cacoullos and Papathanasiou [2]}$$

$$= E[(\mu - pX - \sigma^2 w(X))\Delta f(X)]$$
and, by (2.3) and (2.4), we have
\[ d_{TV}(BB(\alpha, \beta, n), B(n, p)) \leq \frac{(1 - p^{n+1} - q^{n+1})}{(n+1)pq} E|\mu - pX - \sigma^2 w(X)|. \] (2.6)

Using Proposition 2.1, we get
\[
E|\mu - pX - \sigma^2 w(X)| = \sum_{x=0}^{n} |\mu - px - \sigma^2 w(x)| p(x)
\]
\[
= \sum_{x=0}^{n} \left| \frac{n\alpha}{\alpha + \beta} - \frac{\alpha x}{\alpha + \beta} - \frac{(n - x)(\alpha + x)}{\alpha + \beta} \right| p(x)
\]
\[
= \sum_{x=0}^{n} \frac{(n - x)x}{\alpha + \beta} p(x)
\]
\[
= \frac{n^2\alpha}{(\alpha + \beta)^2} - \frac{n\alpha[n(1 + \alpha) + \beta]}{(\alpha + \beta)^2(1 + \alpha + \beta)}
\]
\[
= \frac{n(n - 1)pq}{1 + \alpha + \beta} \] (2.7)

Hence, by (2.6) and (2.7), the theorem is proved. \( \square \)

**Remark.** Note that the result gives a good binomial approximation if \( \frac{\alpha}{\alpha + \beta} \) or \( \frac{n}{\beta} \) is small.

**References**


**Received: March 31, 2008**