Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 11, Number 4 (2015), pp.1967-1970

http://www.ripublication.com

Binomial Approximation for A Sum of Independent Hypergeometric Random Variables

K. Teerapabolarn

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

Abstract

We use Stein's method and the hypergeometric w-functions to determine a bound for the total variation distance between the distribution of a sum of n independent hypergeometric random variables, each with parameters N_i, r_i and m_i, and a binomial distribution with parameters $\sum_{i=1}^{n} m_i$ and $\frac{1}{s}\sum_{i=1}^{n} \frac{r_i m_i}{N_i}$. The result obtained in this study gives a good approximation when N_i is large with respect to r_i and m_i.

AMS subject classification: 60F05, 60G50.
Keywords: Binomial approximation, hypergeometric distribution, hypergeometric w-function, Stein's method.

1. Introduction

Let $X_1, \ldots, X_n$ be independently distributed hypergeometric random variables, each with the probability $P(X_i = k) = \frac{\binom{r_i}{k}\binom{N_i - r_i}{m_i - k}}{\binom{N_i}{m_i}}$ for $k = 0, \ldots, m_i$ $(m_i \leq r_i)$, mean $\mu_i = \frac{r_i m_i}{N_i}$ and variance $\sigma_i^2 = \frac{r_i m_i (N_i - r_i)(N_i - m_i)}{N_i^2 (N_i - 1)}$. Let $\mathcal{S}_n = \sum_{i=1}^{n} X_i$ and $\mathcal{B}_{s,p}$ denote the binomial random variable with parameters $s = \sum_{i=1}^{n} m_i$ and $p = \frac{1}{s}\sum_{i=1}^{n} \mu_i$. For $n = 1$, Teerapabolarn [4] gave a bound for the total variation distance between a

hypergeometric distribution with parameters N, r and m and a binomial distribution with parameters m and $p = \frac{r}{N}$ as follows:

$$d(\mathcal{H}_{N,r,m}, \mathcal{B}_{m,p}) \leq \frac{(1 - p^{m+1} - q^{m+1})r(r-1)}{(m+1)(N-1)}, \tag{1.1}$$

where

$$d(\mathcal{H}_{N,r,m}, \mathcal{B}_{m,p}) = \sup_{A \subseteq \{0,\ldots,m\}} |P(\mathcal{H}_{N,r,m} \in A) - P(\mathcal{B}_{m,p} \in A)|$$

and $\mathcal{H}_{N,r,m}$ is the hypergeometric random variable with parameters N, r and m.

In this paper, we are interested to determine a bound for approximating the distribution of a sum of $n(> 1)$ independent hypergeometric random variables by a binomial distribution with parameters s and p, in the form of $d(\mathcal{S}_n, \mathcal{B}_{s,p})$. The tools for giving the desired result are Stein's method and the hypergeometric w-functions, which are in Section 2. In Section 3, our result is derived by these tools and the conclusion of this study is presented in the last section.

2. Method

The following lemma gives the hypergeometric w-functions [4].

Lemma 2.1. For $1 \leq i \leq n$, let w_i be the w-function associated with the hypergeometric random variable X_i, then we have the following:

$$w_i(k) = \frac{(r_i - k)(m_i - k)}{N\sigma_i^2}, \ k = 0, \ldots, m_i. \tag{2.1}$$

The following relation is an important property for proving the result, which was stated by [2].

$$\begin{aligned} Cov(\mathcal{S}_n, f(\mathcal{S}_n)) &= \sum_{i=1}^{n} Cov\left(X_i, f\left(X_i + \sum_{j \neq i} X_j\right)\right) \\ &= \sum_{i=1}^{n} \sigma_i^2 E[w_i(X_i)\Delta f(\mathcal{S}_n)], \end{aligned} \tag{2.2}$$

for any function $f : \mathbb{N} \cup \{0\} \to \mathbb{R}$ for which $E|w_i(X_i)\Delta f(\mathcal{S}_n)| < \infty$, where $\Delta f(x) = f(x+1) - f(x)$.

For Stein's method in the binomial approximation, it can be applied for every $s \in \mathbb{N}$ and $0 < p = 1 - q < 1$, for every $A \subseteq \{0, \ldots, s\}$ and bounded real-valued function $f = f_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ defined as in [1], where $f(0) = f(1)$ and $f(x) = f(s)$ for $x \geq s$. So, Stein's equation for these conditions is as follows:

$$P(\mathcal{S}_n \in A) - P(\mathcal{B}_{s,p} \in A) = E[(s - \mathcal{S}_n)pf(\mathcal{S}_n + 1) - q\mathcal{S}_n f(\mathcal{S}_n)]. \tag{2.3}$$

For $A \subseteq \{0, \ldots, s\}$ and $x \in \mathbb{N} \cup \{0\}$, Ehm [3] showed that

$$\sup_{A,x} |\Delta f(x)| \leq \frac{1 - p^{s+1} + q^{s+1}}{(s+1)pq}. \tag{2.4}$$

3. Result

The following theorem presents the main result of this study.

Theorem 3.1. With the above definitions, we have the following:

$$d(\mathcal{S}_n, \mathcal{B}_{s,p}) \leq \frac{1 - p^{s+1} + q^{s+1}}{(s+1)pq} \sum_{i=1}^{n} \left\{ |r_i + m_i - pN_i| + \frac{(r_i - 1)(m_i - 1}{N_i - 1} \right\} \frac{r_i m_i}{N_i^2}. \tag{3.1}$$

Proof. From (2.3), it follows that

$$\begin{aligned} d(\mathcal{S}_n, \mathcal{B}_{s,p}) &= |E[(s - \mathcal{S}_n)pf(\mathcal{S}_n + 1) - q\mathcal{S}_n f(\mathcal{S}_n)]| \\ &= |E[spf(\mathcal{S}_n + 1) - p\mathcal{S}_n \Delta f(\mathcal{S}_n) - \mathcal{S}_n f(\mathcal{S}_n)]| \\ &= |E[sp\Delta f(\mathcal{S}_n)] - pE[\mathcal{S}_n \Delta f(\mathcal{S}_n)] - Cov(\mathcal{S}_n, f(\mathcal{S}_n))| \\ &= \left| \sum_{i=1}^{n} \{E[\mu_i \Delta f(\mathcal{S}_n)] - pE[X_i \Delta f(\mathcal{S}_n)] - Cov(X_i, f(\mathcal{S}_n))\} \right|. \end{aligned}$$

Using (2.2) and Lemma 1, we have

$$\begin{aligned} d_{TV}(\mathcal{S}_n, \mathcal{B}_{s,p}) &= \left| \sum_{i=1}^{n} \left\{ E\left[(\mu_i - pX_i)\Delta f(\mathcal{S}_n)\right] - \sigma_i^2 E[w_i(X_i)\Delta f(\mathcal{S}_n)] \right\} \right| \\ &\leq \sum_{i=1}^{n} E\{|\mu_i - pX_i - \sigma_i^2 w_i(X_i)||\Delta f(\mathcal{S}_n)|\} \\ &\leq \sup_{A,x} |\Delta f(x)| \sum_{i=1}^{n} E\left\{ \left| \frac{r_i m_i}{N_i} - pX_i - \frac{(r_i - X_i)(m_i - X_i)}{N_i} \right| \right\} \\ &= \sup_{A,x} |\Delta f(x)| \sum_{i=1}^{n} E\left\{ \left| -pX_i + \frac{(m_i + r_i)X_i}{N_i} - \frac{X_i^2}{N_i} \right| \right\} \\ &\leq \sup_{A,x} |\Delta f(x)| \sum_{i=1}^{n} \left\{ |r_i + m_i - pN_i)|\, E(X_i) - E(X_i^2) \right\} \frac{1}{N_i} \\ &\leq \sup_{A,x} |\Delta f(x)| \sum_{i=1}^{n} \left\{ |r_i + m_i - pN_i| + \frac{(r_i - 1)(m_i - 1)}{N_i - 1} \right\} \frac{r_i m_i}{N_i^2}. \end{aligned}$$

Hence, by (2.4), (3.1) is obtained. ■

4. Conclusion

In this study, a bound on the error of binomial approximation to the distribution of a sum of n independent hypergeometric random variables was obtained, by using Stein's method and the hypergeometric w-functions. The distribution of this summands can be approximated by a binomial distribution with parameters $s = \sum_{i=1}^{n} m_i$ and $p = \frac{1}{s}\sum_{i=1}^{n} \frac{r_i m_i}{N_i}$ when for each $i \in \{1, \ldots, n\}$, N_i is large with respect to r_i and m_i.

References

[1] Barbour, A. D., Holst, L., Janson, S., 1992, "Poisson approximation", Oxford Studies in Probability 2, Clarendon Press, Oxford.

[2] Cacoullos, T., Papathanasiou, V., 1989, "Characterization of distributions by variance bounds", Statist. Probab. Lett., 7(5), 351–356.

[3] Ehm, W., 1991, "Binomial approximation to the Poisson binomial distribution", Statist. Probab. Lett., 11(1), 7–16.

[4] Wongkasem, P., Teerapabolarn, K., Gulasirima, R., 2008, "On approximating a generalized binomial by binomial and Poisson distributions", Int. J. Statist. Systems, 3(2), 113-–124.