International Journal of Pure and Applied Mathematics
Volume 83 No. 4 2013, 559-563
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: http://dx.doi.org/10.12732/ijpam.v83i4.4

BINOMIAL APPROXIMATION TO THE GENERALIZED HYPERGEOMETRIC DISTRIBUTION

K. Teerapabolarn

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

Abstract: In this paper, we use Stein's method and the w-function associated with the generalized hypergeometric random variable to give a bound for the total variation distance between the binomial and generalized hypergeometric distributions.

AMS Subject Classification: 62E17, 60F05
Key Words: binomial approximation, generalized hypergeometric distribution, total variation distance, Stein's method

1. Introduction

The discrete random variable X is said to be the generalized hypergeometric random variable if its probability mass function is as follows: (Crosu[3])

$$p_X(x) = \binom{N-1}{x}\frac{\Gamma(N+\alpha-x)\Gamma(\beta+1+x)\Gamma(\alpha+\beta+2)}{\Gamma(\alpha+1)\Gamma(\beta+1)\Gamma(\alpha+\beta+N+1)},\ x = 0, ..., N-1,$$

where $N \in \mathbb{N}\setminus\{1\}, \alpha \geq 0$ and $\beta > -1$ and the mean and variance of X are

$$\mu = \frac{(N-1)(\beta+1)}{\alpha+\beta+2}$$

and

Received: November 5, 2012

© 2013 Academic Publications, Ltd.
url: www.acadpubl.eu

$$\sigma^2 = \frac{(N-1)(\beta+1)(\alpha+1)(\alpha+\beta+N+1)}{(\alpha+\beta+2)^2(\alpha+\beta+3)},$$

respectively. Let $\mathbb{GH}_{\alpha,\beta,N}$ be the generalized hypergeometric distribution with parameters α, β and N. It is well-known that the hypergeometric distribution can be approximated by the binomial and Poisson distributions if some conditions of their parameters are satisfied. Similarly, following this fact, each of the binomial and Poisson distributions can also be used as an approximation of the generalized hypergeometric distribution. In this case, Crosu[3] used the Stein-Chen method and the w-function associated with the generalized hypergeometric random variable to obtain a bound for the total variation distance between the generalized hypergeometric and Poisson distributions, when $\beta + 2 \geq N$, as follows:

$$\begin{aligned} d_{TV}(\mathbb{GH}_{\alpha,\beta,N}, \mathbb{P}_\mu) &= \sup_{A \subseteq \mathbb{N}\cup\{0\}} |\mathbb{GH}_{\alpha,\beta,N}\{A\} - \mathbb{P}_\mu\{A\}| \\ &\leq (1-e^{-\mu})\frac{(\alpha+1)(\beta+3-N)+(\beta+1)(\beta+2)}{(\alpha+\beta+2)(\alpha+\beta+3)}, \end{aligned} \quad (1.1)$$

where $\mathbb{P}_\mu$ is a Poisson distribution with mean $\mu = \frac{(N-1)(\beta+1)}{\alpha+\beta+2}$.

In this paper, we are interested to determine a bound for the total variation distance between the generalized hypergeometric and binomial distributions,

$$d_{TV}(\mathbb{GH}_{\alpha,\beta,N}, \mathbb{B}_{n,p}) = \sup_{A\subseteq\{0,\ldots,n\}} |\mathbb{GH}_{\alpha,\beta,N}\{A\} - \mathbb{B}_{n,p}\{A\}|, \quad (1.2)$$

where $\mathbb{B}_{n,p}$ is a binomial distribution with parameters $n = N-1$ and $p = \frac{\beta+1}{\alpha+\beta+2}$.

2. Method

The tools for deriving the result for this approximation consist of Stein's method for the binomial distribution and the w-function associated with the generalized hypergeometric random variable. Following [1], Stein's equation for binomial distribution with parameters $n \in \mathbb{N}$ and $p \in (0,1)$ is, for given h, of the form

$$h(x) - \mathcal{B}_{n,p}(h) = (n-x)pg(x+1) - qxg(x), \quad (2.1)$$

where $\mathcal{B}_{n,p}(h) = \sum_{k=0}^{n} h(k)\binom{n}{k}p^k q^{n-k}$ and g and h are bounded real-valued functions defined on $\{0,1,...,n\}$.

For $A \subseteq \{0,1,...,n\}$, let $h_A : \{0,1,...,n\} \to \mathbb{R}$ be defined by

$$h_A(x) = \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases} \quad (2.2)$$

Following [1], let $g_A : \mathbb{N} \cup \{0\} \to \mathbb{R}$ satisfy (2.1), where $g_A(0) = g_A(1)$ and $g_A(x) = g_A(n)$ for $x \geq n$. Let $x \in \mathbb{N}$ and $\Delta g_A(x)$=$g_A(x+1) - g_A(x)$, Ehm [4] showed that

$$\sup_A |\Delta g_A(x)| \leq \frac{1 - p^{n+1} - q^{n+1}}{(n+1)pq}. \tag{2.3}$$

For the w-function associated with the generalized hypergeometric random variable, following [2], if $E|w(X)\Delta g_A(X)| < \infty$ then

$$Cov(X, g_A(X)) = \sigma^2 E[w(X)\Delta g_A(X)] \tag{2.4}$$

and we have $E[w(X)] = 1$. The following lemma presents this w-function, which obtained from [3].

Lemma 2.1. *Let $w(X)$ be the w-function associated with the generalized hypergeometric random variable X. Then, we have the following:*

$$w(x) = \frac{(\beta + x + 1)(N - x - 1)}{(\alpha + \beta + 2)\sigma^2} \ , \ x = 0, ..., N-1. \tag{2.5}$$

3. Result

The following theorem presents a bound for the total variation distance between $\mathbb{GH}_{\alpha,\beta,N}$ and $\mathbb{B}_{n,p}$.

Theorem 3.1. *For $n = N - 1$ and $p = 1 - q = \frac{\beta+1}{\alpha+\beta+2}$. Then we have*

$$d_{TV}(\mathbb{GH}_{\alpha,\beta,N}, \mathbb{B}_{n,p}) \leq \left(1 - p^{n+1} - q^{n+1}\right) \frac{(N-1)(N-2)}{N(\alpha + \beta + 3)}. \tag{3.1}$$

Proof. For $A \subseteq \{0, 1, ..., n\}$, substituting h, x by h_A, X respectively and taking expectation in (2.1), we obtain

$$\mathbb{GH}_{\alpha,\beta,N}\{A\} - \mathbb{B}_{n,p}\{A\} = E[(n - X)pg(X+1) - qXg(X)], \tag{3.2}$$

where $g = g_A$ is defined as mentioned above.

Let $\delta(\mathbb{GH}, \mathbb{B}) = \mathbb{GH}_{\alpha,\beta,N}\{A\} - \mathbb{B}_{n,p}\{A\}$, then we obtain

$$\begin{aligned} \delta(\mathbb{GH}, \mathbb{B}) &= E[npg(X+1) - pX\Delta g(X) - Xg(X)] \\ &= E[npg(X+1)] - pE[X\Delta g(X)] - E[Xg(X)] \end{aligned}$$

$$= npE[g(X+1)] - pE[X\Delta g(X)] - Cov(X, g(X)) - \mu E[g(X)]$$
$$= npE[\Delta g(X)] - pE[X\Delta g(X)] - Cov(X, g(X)).$$

Using Lemma 2.1 and (2.4), we have

$$E|w(X)\Delta g(X)| \leq \frac{1-p^{n+1}-q^{n+1}}{(n+1)pq} E|w(X)| = \frac{1-p^{n+1}-q^{n+1}}{(n+1)pq} E[w(X)] < \infty.$$

From which it follows that

$$\begin{aligned}
\delta(\mathbb{GH}, \mathbb{B}) &= npE[\Delta g(X)] - pE[X\Delta g(X)] - \sigma^2 E[w(X)\Delta g(X)] \\
&= E\{[(n-X)p - \sigma^2 w(X)]\Delta g(X)\} \\
&= E\left\{\left[\frac{(N-1-X)(\beta+1)}{\alpha+\beta+2} - \frac{(\beta+X+1)(N-X-1)}{(\alpha+\beta+2)}\right]\Delta g(X)\right\} \\
&= E\left\{\left[-\frac{X(N-X-1)}{(\alpha+\beta+2)}\right]\Delta g(X)\right\}.
\end{aligned}$$

Therefore, it follows from (3.2) and (1.2), we obtain

$$\begin{aligned}
d_{TV}(\mathbb{GH}_{\alpha,\beta,N}, \mathbb{B}_{n,p}) &\leq E\{|(n-X)p - \sigma^2 w(X)||\Delta g(X)|\} \\
&= \frac{1-p^{n+1}-q^{n+1}}{(n+1)pq} E\{\sigma^2 w(X) - (n-X)p]\} \\
&= \frac{1-p^{n+1}-q^{n+1}}{(n+1)pq} E\{\sigma^2 - (1-p)\mu\} \\
&= \left(1-p^{n+1}-q^{n+1}\right)\frac{(N-1)(N-2)}{N(\alpha+\beta+3)}.
\end{aligned}$$

Hence, we have the theorem. □

4. Conclusion

In the present study, a bound for the total variation distance between the generalized hypergeometric and binomial distributions is obtained by using Stein's method and the w-function associated with the generalized hypergeometric random variable. With this bound, it is observed that if $\frac{N}{\alpha+\beta}$ is small, or $\alpha+\beta$ is large, then the result in Theorem 3.1 gives a good binomial approximation, that is, the binomial distribution with parameters $n = N-1$ and $p = \frac{\beta+1}{\alpha+\beta+2}$ can be used as an approximation of the generalized hypergeometric distribution with parameters α, β and N when α or β is sufficiently large. Moreover, it has no any conditions of their parameters in this approximation, which is different from the Poisson approximation as mentioned in (1.1).

References

[1] A.D. Barbour, L. Holst, S. Janson, *Poisson Approximation*, Oxford Studies in Probability 2, Clarendon Press, Oxford (1992).

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

[3] C. Crosu, Some application of Chernoff bounds, *Scientific Journal of the Technical University of Civil Engineering*, **6** (2010), 66-71.

[4] W. Ehm, Binomial approximation to the Poisson binomial distribution, *Statist. Probab. Lett.*, **11** (1991), 7-16.