International Journal of Pure and Applied Mathematics
Volume 90 No. 4 2014, 515-518
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: http://dx.doi.org/10.12732/ijpam.v90i4.12

ijpam.eu

AN IMPROVED BINOMIAL TO APPROXIMATE THE HYPERGEOMETRIC DISTRIBUTION

K. Teerapabolarn
Department of Mathematics
Faculty of Science
Burapha University
Chonburi, 20131, THAILAND

Abstract: We give an improved binomial distribution with parameters n and $p = \frac{m}{N}$ for approximating the hypergeometric distribution with parameters N, m and n. The improved approximation is more accurate than the binomial approximation when N is sufficiently large.

AMS Subject Classification: 62E17, 60F05
Key Words: binomial approximation, binomial probability function, hypergeometric probability function

1. Introduction

Let X be the hypergeometric random variable with parameters N, m and n, and its probability function is of the form

$$\mathbf{h}_{N,m,n}(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}},\ x = 0, 1, ..., n, \tag{1.1}$$

where $N, m, n \in \mathbb{N}$ and $N > m \geq n$. The mean and variance of X are $E(X) = \frac{mn}{N}$ and $Var(X) = \frac{mn(N-m)(N-n)}{(N-1)N^2}$, respectively. This distribution is obtained from a scheme of sampling without replacement with sample size n. It is well-known that, if $N \to \infty$ while $\frac{m}{N}$ remains a constant, then $\mathbf{h}_{N,m,n}(x) \to \mathbf{b}_{n,p}(x) =$

Received: November 22, 2013

$\binom{n}{x}p^x q^{n-x}$ for every $x \in \{0, ..., n\}$, where $p = 1-q = \frac{m}{N}$. Therefore, the binomial probability function can be used as an approximation of the hypergeometric probability function if N is large. In this case, Teerapabolarn and Wongkasem [2] gave a pointwise bound on $|\mathbf{h}_{N,m,n}(x) - \mathbf{b}_{n,p}(x)|$ for $x \in \{0, ..., n\}$.

In this paper, we give an improved binomial probability function, $\widehat{\mathbf{b}}_{n,p}(x)$, for approximating the hypergeometric probability function, and the accuracy of the approximation is measured in the form of $\left|\mathbf{h}_{N,m,n}(x) - \widehat{\mathbf{b}}_{n,p}(x)\right|$ for $x \in \{0, ..., n\}$. The result of this study is in Section 2. In Section 3, some numerical examples have been given to illustrate the improved approximation and the conclusion is presented in the last section.

2. Result

The following lemma is also directly obtained from [1].

Lemma 2.1. *For $x, N \in \mathbb{N}$, then*

$$\prod_{i=0}^{x-1}\left(p - \frac{i}{N}\right) = p^x\left[1 - \frac{x(x-1)}{2Np}\right] + O\left(\frac{1}{N^2}\right), \tag{2.1}$$

$$\frac{1}{\prod_{i=0}^{x-1}\left(1 - \frac{i}{N}\right)} = 1 + \frac{x(x-1)}{2N} + O\left(\frac{1}{N^2}\right). \tag{2.2}$$

Theorem 2.1. *Let $x \in \{0, ..., n\}$ and $p = \frac{m}{N}$. Then we have the following:*

$$\mathbf{h}_{N,m,n}(x) = \widehat{\mathbf{b}}_{n,p}(x) + O\left(\frac{1}{N^2}\right) \tag{2.3}$$

and for large N,

$$\widehat{\mathbf{b}}_{n,p}(x) = \mathbf{h}_{N,m,n}(x), \tag{2.4}$$

where $\widehat{\mathbf{b}}_{n,p}(x) = \mathbf{b}_{n,p}(x)\left\{1 + \frac{n(n-1)}{2N} - \frac{(n-x)(n-x-1)}{2(N-m)}\right\} / \left\{1 + \frac{x(x-1)}{2m}\right\}$.

Proof. For $x = 0$, applying Lemma 2.1, we have that

$$\begin{aligned}\mathbf{h}_{N,m,n}(0) &= \frac{(N-m)\cdots(N-m-n+1)}{(N)\cdots(N-n+1)}\\ &= \frac{\prod_{i=0}^{n-1}\left(q - \frac{i}{N}\right)}{\prod_{i=0}^{n-1}\left(1 - \frac{i}{N}\right)}\end{aligned}$$

$$= q^n \left\{1 + \frac{n(n-1)}{2N} - \frac{n(n-1)}{2Nq}\right\} + O\left(\frac{1}{N^2}\right)$$
$$= \widehat{\mathbf{b}}_{n,p}(0) + O\left(\frac{1}{N^2}\right).$$

Next, we have to show that (2.3) holds for $x \in \{1, ..., n\}$. Using Lemma 2.1, we can obtain

$$\begin{aligned}\mathbf{h}_{N,m,n}(x) &= \binom{n}{x} \frac{m!(N-n)!(N-m)!}{(m-x)!N![(N-m)-(n-x)]!} \\ &= \binom{n}{x} \frac{[m\cdots(m-x+1)][(N-m)\cdots(N-m-n+x+1)]}{N\cdots(N-n+1)} \\ &= \binom{n}{x} \frac{\prod_{i=0}^{x-1}\left(p-\frac{i}{N}\right)\prod_{i=0}^{n-x-1}\left(q-\frac{i}{N}\right)}{\prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right)} \\ &= \binom{n}{x} \frac{p^x q^{n-x}}{1+\frac{x(x-1)}{2Np}} \left\{1 + \frac{n(n-1)}{2N} - \frac{(n-x)(n-x-1)}{2Nq}\right\} + O\left(\frac{1}{N^2}\right) \\ &= \widehat{\mathbf{b}}_{n,p}(x) + O\left(\frac{1}{N^2}\right).\end{aligned}$$

As N is large, we get $O\left(\frac{1}{N^2}\right) = 0$. Hence $\widehat{\mathbf{b}}_{n,p}(x) = \mathbf{h}_{N,m,n}(x)$. □

3. Numerical Examples

The following examples have been given to illustrate how well the improved binomial distribution approximates the hypergeometric distribution (when N is sufficiently large or $\frac{n}{N}$ is small).

3.1. Let $N = 100$, $m = 20$ and $n = 10$, then $p = 0.2$ and the numerical results are as follows:

x	$\mathbf{h}_{N,m,n}(x)$	$\widehat{\mathbf{b}}_{n,p}(x)$	$\mathbf{b}_{n,p}(x)$	$\lvert\mathbf{h}_{N,m,n}(x) - \widehat{\mathbf{b}}_{n,p}(x)\rvert$	$\lvert\mathbf{h}_{N,m,n}(x) - \mathbf{b}_{n,p}(x)\rvert$
0	0.09511627	0.09529459	0.10737418	0.00017831	0.01225791
1	0.26793316	0.26843546	0.26843546	0.00050229	0.00050229
2	0.31817063	0.31637036	0.30198989	0.00180027	0.01618074
3	0.20920809	0.20789159	0.20132659	0.00131650	0.00788149
4	0.08410730	0.08553960	0.08808038	0.00143230	0.00397308
5	0.02153147	0.02334130	0.02642412	0.00180983	0.00489265
6	0.00354136	0.00432538	0.00550502	0.00078402	0.00196366
7	0.00036793	0.00054187	0.00078643	0.00017394	0.00041850
8	0.00002300	0.00004416	0.00007373	0.00002116	0.00005073
9	0.00000078	0.00000212	0.00000410	0.00000134	0.00000332

3.2. Let $N = 500$, $m = 50$ and $n = 30$, then $p = 0.1$ and the numerical results are as follows:

x	$\mathbf{h}_{N,m,n}(x)$	$\widehat{\mathbf{b}}_{n,p}(x)$	$\mathbf{b}_{n,p}(x)$	$\lvert\mathbf{h}_{N,m,n}(x) - \widehat{\mathbf{b}}_{n,p}(x)\rvert$	$\lvert\mathbf{h}_{N,m,n}(x) - \mathbf{b}_{n,p}(x)\rvert$
0	0.03832346	0.03829335	0.04239116	0.00003011	0.00406770
1	0.13654439	0.13675074	0.14130386	0.00020634	0.00475947
2	0.22989287	0.22988814	0.22765622	0.00000473	0.00223665
3	0.24347992	0.24276967	0.23608793	0.00071025	0.00739199
4	0.18217926	0.18145747	0.17706595	0.00072179	0.00511331
5	0.10253477	0.10258895	0.10230477	0.00005418	0.00023000
6	0.04512974	0.04578454	0.04736332	0.00065480	0.00223358
7	0.01594413	0.01661722	0.01804317	0.00067309	0.00209904
8	0.00460536	0.00501253	0.00576379	0.00040717	0.00115843
9	0.00110214	0.00127726	0.00156547	0.00017512	0.00046334
10	0.00022068	0.00027834	0.00036528	0.00005765	0.00014459
11	0.00003724	0.00005236	0.00007379	0.00001512	0.00003656
12	0.00000532	0.00000856	0.00001298	0.00000324	0.00000766

The improved approximation is better than the binomial approximation.

4. Conclusion

The result of this study is an improved binomial distribution with parameters n and $p = \frac{m}{N}$. It is more accurate for approximating the hypergeometric distribution when N is sufficiently large, and also the improved approximation is better than the binomial approximation.

References

[1] D.P. Hu, Y.Q. Cui, A.H. Yin , An improved negative binomial approximation for negative hypergeometric distribution, *Appl. Mech. Mater.*, **427-429** (2013), 2549–2553.

[2] K. Teerapabolarn, P. Wongkasem, A pointwise binomial approximation by w-functions, *Int. J. Pure Appl. Math.*, **71** (2011), 57–66.