International Journal of Pure and Applied Mathematics
Volume 91 No. 2 2014, 231-235
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: http://dx.doi.org/10.12732/ijpam.v91i2.9

ijpam.eu

AN IMPROVED NEGATIVE BINOMIAL DISTRIBUTION TO APPROXIMATE THE NEGATIVE HYPERGEOMETRIC DISTRIBUTION

K. Teerapabolarn

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

Abstract: We give an improved negative binomial distribution with parameters r and p for approximating the negative hypergeometric distribution with parameters R, S and r, where $p = 1 - q = \frac{R+1}{R+S+1}$. The improved approximation is more accurate than the negative binomial approximation when R is sufficiently large.

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

1. Introduction

Let X be the negative hypergeometric random variable with parameters R, S and r, and its probability function in our attention is of the form

$$\mathbf{nh}_{R,S,r}(x) = \frac{\binom{r+x-1}{x}\binom{R-r+S-x}{S-x}}{\binom{R+S}{S}},\ x = 0, 1, ..., S, \tag{1.1}$$

where $R, S \in \mathbb{N}$ and $r \in \{1, ..., R\}$. The mean and variance of X are $E(X) = \frac{rS}{R+1}$ and $Var(X) = \frac{rS(R+S+1)(R-r+1)}{(R+1)^2(R+2)}$, respectively. This distribution was used

Received: November 22, 2013

by Kaigh and Lachenbruch [2] in resampling for nonparametric quantile estimation. More references of this distribution can be found in [4]. Note that this distribution is obtained from a finite sample analogy to the negative binomial distribution, which arises in a scheme of sampling with replacement. In addition, if $R, S \to \infty$ while $\frac{S}{R+1}$ remains a constant, then $\mathbf{nh}_{R,S,r}(x) \to \mathbf{nb}_{r,p}(x) = \binom{r+x-1}{x} p^r q^x$ for every $x \in \{0, ..., S\}$, where $p = 1 - q = \frac{R+1}{R+S+1}$ [4]. Therefore, the negative binomial probability function can be used as an estimate of the negative hypergeometric probability function when R is large. In this case, Malingam and Teerapabolarn [3] gave a bound on $|\mathbf{nh}_{R,S,r}(x) - \mathbf{nb}_{r,p}(x)|$ for $x \in \{0, ..., S\}$.

In this paper, we focus on determining an improved negative binomial probability function, $\widehat{\mathbf{nb}}_{r,p}(x)$, for approximating the negative hypergeometric probability function, and the accuracy of the approximation is measured in the form of $\left|\mathbf{nh}_{R,S,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x)\right|$ for $x \in \{0, ..., S\}$. The result of this study is in Section 2. In Section 3, some numerical examples are given to illustrate the improved approximation and the conclusion of this study is presented in the last section.

2. Result

The following lemma directly follows from [1].

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

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

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

Theorem 2.1. *Let $x \in \{0, ..., S\}$, $p = \frac{R+1}{R+S+1}$. Then we have the following:*

$$\mathbf{nh}_{R,S,r}(x) = \widehat{\mathbf{nb}}_{r,p}(x) + O\left(\frac{1}{(R+S+1)^2}\right) \tag{2.3}$$

and for large R,

$$\widehat{\mathbf{nb}}_{r,p}(x) = \mathbf{nh}_{R,S,r}(x), \tag{2.4}$$

where $\widehat{\mathbf{nb}}_{r,p}(x) = \mathbf{nb}_{r,p}(x)\left\{1 + \frac{(r+x+1)(r+x)}{2(R+S+1)} - \frac{(r+1)r}{2(R+1)}\right\} / \left\{1 + \frac{x(x-1)}{2S}\right\}$.

Proof. For $x = 0$, applying Lemma 2.1, it follows that

$$\begin{aligned}
\mathbf{nh}_{R,S,r}(0) &= \frac{(R+S-r)!R!}{(R+S)!(R-r)!} \\
&= \frac{R\cdots(R-(r-1))}{(R+S)\cdots(R+S-(r-1))} \\
&= \frac{\prod_{i=1}^{r}\left(p-\frac{i}{R+S+1}\right)}{\prod_{i=1}^{r}\left(1-\frac{i}{R+S+1}\right)} \\
&= p^r\left\{1+\frac{(r+1)r}{2(R+S+1)}-\frac{(r+1)r}{2(R+S+1)p}\right\}+O\left(\frac{1}{(R+S+1)^2}\right) \\
&= \mathbf{nb}_{r,p}(0)\left\{1-\frac{q(r+1)r}{2(R+S+1)p}\right\}+O\left(\frac{1}{(R+S+1)^2}\right) \\
&= \widehat{\mathbf{nb}}_{r,p}(0)+O\left(\frac{1}{(R+S+1)^2}\right).
\end{aligned}$$

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

$$\begin{aligned}
\mathbf{nh}_{R,S,r}(x) &= \binom{r+x-1}{x}\frac{(R+S-r-x)!R!S!}{(R+S)!(R-r)!(S-x)!} \\
&= \binom{r+x-1}{x}\frac{[R\cdots(R-(r-1))][S\cdots(S-(x-1))]}{(R+S)\cdots(R+S-(r+x-1))} \\
&= \binom{r+x-1}{x}\frac{\prod_{i=1}^{r}\left(p-\frac{i}{R+S+1}\right)\prod_{i=0}^{x-1}\left(q-\frac{i}{R+S+1}\right)}{\prod_{i=1}^{r+x}\left(1-\frac{i}{R+S+1}\right)} \\
&= \binom{r+x-1}{x}\frac{p^r q^x}{1+\frac{x(x-1)}{2(R+S+1)q}}\left\{1+\frac{(r+x+1)(r+x)}{2(R+S+1)}\right. \\
&\qquad \left.-\frac{(r+1)r}{2(R+S+1)p}\right\}+O\left(\frac{1}{(R+S+1)^2}\right) \\
&= \frac{\mathbf{nb}_{r,p}(x)}{1+\frac{x(x-1)}{2S}}\left\{1+\frac{(r+x+1)(r+x)}{2(R+S+1)}-\frac{(r+1)r}{2(R+1)}\right\} \\
&\qquad +O\left(\frac{1}{(R+S+1)^2}\right) \\
&= \widehat{\mathbf{nb}}_{r,p}(x)+O\left(\frac{1}{(R+S+1)^2}\right).
\end{aligned}$$

Also, if R is large, then $O\left(\frac{1}{(R+S+1)^2}\right) = 0$. Hence $\widehat{\mathbf{nb}}_{r,p}(x) = \mathbf{nh}_{R,S,r}(x)$. □

3. Numerical Examples

The following examples are given to illustrate how well the improved negative binomial distribution approximates the negative hypergeometric distribution (when R is sufficiently large).

3.1. Let $R = 50$, $S = 10$ and $r = 5$, then $p = \frac{51}{61}$ and the numerical results are as follows:

x	$\mathbf{nh}_{R,S,r}(x)$	$\widehat{\mathbf{nb}}_{r,p}(x)$	$\mathbf{nb}_{r,p}(x)$	$\lvert\mathbf{nh}_{R,S,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x)\rvert$	$\lvert\mathbf{nh}_{R,S,r}(x) - \mathbf{nb}_{r,p}(x)\rvert$
0	0.38794385	0.38881240	0.40850907	0.00086854	0.02056522
1	0.35267623	0.35163411	0.33484350	0.00104212	0.01783273
2	0.17633811	0.17439289	0.16467713	0.00194522	0.01166098
3	0.06210651	0.06279968	0.06299125	0.00069317	0.00088475
4	0.01672098	0.01863389	0.02065287	0.00191291	0.00393189
5	0.00354091	0.00489835	0.00609429	0.00135744	0.00255338
6	0.00059015	0.00119078	0.00166511	0.00060063	0.00107495
7	0.00007570	0.00027461	0.00042895	0.00019890	0.00035325
8	0.00000710	0.00006100	0.00010548	0.00005391	0.00009838
9	0.00000044	0.00001318	0.00002498	0.00001274	0.00002454
10	0.00000001	0.00000279	0.00000573	0.00000277	0.00000572

3.2. Let $R = 100$, $S = 30$ and $r = 10$, then $p = \frac{101}{131}$ and the numerical results are as follows:

x	$\mathbf{nh}_{R,S,r}(x)$	$\widehat{\mathbf{nb}}_{r,p}(x)$	$\mathbf{nb}_{r,p}(x)$	$\lvert\mathbf{nh}_{R,S,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x)\rvert$	$\lvert\mathbf{nh}_{R,S,r}(x) - \mathbf{nb}_{r,p}(x)\rvert$
0	0.06497833	0.06496122	0.07421655	0.00001711	0.00923822
1	0.16244583	0.16303774	0.16996157	0.00059191	0.00751574
2	0.21773201	0.21770582	0.21407374	0.00002619	0.00365828
3	0.20666089	0.20502982	0.19609808	0.00163108	0.01056282
4	0.15499567	0.15288015	0.14595086	0.00211552	0.00904481
5	0.09727315	0.09626405	0.09358681	0.00100909	0.00368633
6	0.05286584	0.05335211	0.05358024	0.00048627	0.00071440
7	0.02543920	0.02678237	0.02804636	0.00134317	0.00260716
8	0.01100302	0.01243043	0.01364852	0.00142741	0.00264550
9	0.00432261	0.00541535	0.00625123	0.00109273	0.00192861
10	0.00155380	0.00223965	0.00272000	0.00068584	0.00116620
11	0.00051365	0.00088691	0.00113255	0.00037325	0.00061889
12	0.00015669	0.00033853	0.00045388	0.00018184	0.00029720
13	0.00004419	0.00012520	0.00017590	0.00008101	0.00013171
14	0.00001154	0.00004505	0.00006618	0.00003351	0.00005464
15	0.00000279	0.00001582	0.00002425	0.00001304	0.00002146
16	0.00000062	0.00000544	0.00000868	0.00000482	0.00000806
17	0.00000013	0.00000183	0.00000304	0.00000171	0.00000291
18	0.00000002	0.00000061	0.00000104	0.00000044	0.00000102

For approximating the negative hypergeometric distribution in the examples 3.1 and 3.2, it can be seen that the improved negative binomial distribution is better than the negative binomial distribution.

4. Conclusion

In this study, an improved negative binomial distribution with parameters r and $p = \frac{R+1}{R+S+1}$ was obtained by using some mathematical manipulations. It is more appropriate for approximating the negative hypergeometric distribution, that is, the improved negative binomial distribution can be used as an estimate of the negative hypergeometric distribution when R is sufficiently large. In addition, the improvement of the approximation is more accurate than the negative binomial approximation.

References

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

[2] W.D. Kaigh, P.A. Lachenbruch, A generalized quantile estimator, *Communications in Statistics-Theory and Methods*, **19** (1982), 2217–2238.

[3] P. Malingam, K. Teerapabolarn, A pointwise negative binomial approximation by w-functions, *International Journal of Pure and Applied Mathematics*, **69** (2011), 453–467.

[4] K. Teerapabolarn, On the Poisson approximation to the negative hypergeometric distribution, *Bulletin of the Malaysian Mathematical Sciences Society*, **34** (2011), 331–336.