International Journal of Pure and Applied Mathematics
Volume 93 No. 5 2014, 629-632
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
doi: http://dx.doi.org/10.12732/ijpam.v93i5.4
ijpam.eu

AN IMPROVED BINOMIAL DISTRIBUTION TO APPROXIMATE THE PÓLYA DISTRIBUTION

K. Teerapabolarn

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

Abstract: This paper gives an improved binomial distribution with parameters n and p to approximate the Pólya distribution with parameters N, m, r and c, where $p = 1 - q = \frac{r}{N}$. 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, Pólya probability function

1. Introduction

Draw a ball at random from an urn containing r red and $N - r$ black balls, note the color, and return it into the urn together with c additional balls of the same color. Repeat this way for n draws. Let X be the number of red balls taken out in the n drawings, then the distribution of X is a Pólya distribution with parameters N, n, r and c. The probability function of X is given by

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

where $N, n, r, c \in \mathbb{N}$ and the mean and variance of X are $\mu = \frac{nr}{N}$ and $\sigma^2 =$

Received: December 25, 2013

$\frac{rn(N+cn)(N-r)}{N^2(N+c)}$, respectively. In addition, if $N \to \infty$ while $p = 1-q = \frac{r}{N}$ remains a constant, then $\mathbf{py}_{N,n,r,c}(x) \to \mathbf{b}_{n,p}(x) = \binom{n}{x} p^x q^{n-x}$ for every $x = 0, 1, ..., n$ [2], that is, the binomial distribution can be used as an estimate of the Pólya distribution when N is sufficiently large. In this case, Teerapabolarn [2] gave a bound on the total variation distance between two such distributions.

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

2. Result

Applying the property in [1], the following lemma is also obtained.

Lemma 2.1. *For $x, N \in \mathbb{N}$ and $0 < p < 1$, then*

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

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

Theorem 2.1. *Let $x \in \{0, ..., n\}$ and $p = 1 - q = \frac{r}{N}$. Then we have the following:*

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

and for large N,

$$\widehat{\mathbf{b}}_{n,p}(x) \approx \mathbf{py}_{N,n,r,c}(x), \tag{2.4}$$

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

Proof. Let $\delta = \begin{cases} 0 & \text{if } x = 0, \\ x-1 & \text{if } x = 1, ..., n. \end{cases}$ Applying Lemma 2.1, it follows that

$$\mathbf{py}_{N,n,r,c}(x) = \binom{n}{x}\frac{\left[\left(\frac{r}{c} + \delta\right)\cdots\frac{r}{c}\right]\left[\left(\frac{N-r}{c} + n - x - 1\right)\cdots\frac{N-r}{c}\right]}{\left(\frac{N}{c} + n - 1\right)\cdots\frac{N}{c}}$$

$$= \binom{n}{x} \frac{\prod_{i=0}^{\delta}\left(p+\frac{i}{N/c}\right)\prod_{i=0}^{n-x-1}\left(q+\frac{i}{N/c}\right)}{\prod_{i=0}^{n-1}\left(1+\frac{i}{N/c}\right)}$$

$$= \binom{n}{x} \frac{p^x q^{n-x}}{1+\frac{cn(n-1)}{2N}} \left\{1+\frac{c(n-x)(n-x-1)}{2(N-r)}+\frac{cx(x-1)}{2r}\right\} + O\left(\frac{1}{N^2}\right)$$

$$= \frac{\mathbf{b}_{n,p}(x)}{\frac{1}{c}+\frac{n(n-1)}{2N}} \left\{\frac{1}{c}+\frac{(n-x)(n-x-1)}{2(N-r)}+\frac{x(x-1)}{2r}\right\} + O\left(\frac{1}{N^2}\right)$$

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

Also, if N is large, then $O\left(\frac{1}{N^2}\right) \approx 0$. Hence $\widehat{\mathbf{b}}_{n,p}(x) \approx \mathbf{py}_{N,n,r,c}(x)$. □

3. Numerical examples

The following examples are given to illustrate how well the improved binomial distribution approximates the Pólya distribution.

3.1. Let $N = 100$, $n = 10$, $r = 10$ and $c = 1$, then $p = \frac{10}{100}$ and the numerical results are as follows:

x	$\mathbf{py}_{N,n,r,c}(x)$	$\widehat{\mathbf{b}}_{n,p}(x)$	$\mathbf{b}_{n,p}(x)$	$\lvert\mathbf{py}_{N,n,r,c}(x)-\widehat{\mathbf{b}}_{n,p}(x)\rvert$	$\lvert\mathbf{py}_{N,n,r,c}(x)-\mathbf{b}_{n,p}(x)\rvert$
0	0.36541727	0.36070183	0.34867844	0.00471544	0.01673883
1	0.36910835	0.37406116	0.38742049	0.00495281	0.01831214
2	0.18643738	0.18851495	0.19371024	0.00207757	0.00727286
3	0.06150512	0.06069423	0.05739563	0.00081089	0.00410949
4	0.01457543	0.01359756	0.01116026	0.00097787	0.00341517
5	0.00257755	0.00216649	0.00148803	0.00041106	0.00108952
6	0.00034276	0.00024389	0.00013778	0.00009887	0.00020498
7	0.00003370	0.00001890	0.00000875	0.00001479	0.00002495
8	0.00000233	0.00000096	0.00000036	0.00000138	0.00000197
9	0.00000010	0.00000003	0.00000001	0.00000007	0.00000009

3.2. Let $N = 500$, $n = 30$, $r = 30$ and $c = 2$, then $p = \frac{30}{500}$ and the numerical results are as follows:

x	$\mathbf{py}_{N,n,r,c}(x)$	$\widehat{\mathbf{b}}_{n,p}(x)$	$\mathbf{b}_{n,p}(x)$	$\lvert\mathbf{py}_{N,n,r,c}(x) - \widehat{\mathbf{b}}_{n,p}(x)\rvert$	$\lvert\mathbf{py}_{N,n,r,c}(x) - \mathbf{b}_{n,p}(x)\rvert$
0	0.17317310	0.16258931	0.15625561	0.01058379	0.01691749
1	0.29518142	0.29786527	0.29921286	0.00268384	0.00403144
2	0.26038817	0.27037944	0.27693105	0.00999127	0.01654288
3	0.15769055	0.16218740	0.16498020	0.00449685	0.00728965
4	0.07340767	0.07219686	0.07108190	0.00121081	0.00232577
5	0.02789491	0.02534336	0.02359314	0.00255156	0.00430177
6	0.00897520	0.00726973	0.00627477	0.00170547	0.00270043
7	0.00250471	0.00174236	0.00137320	0.00076235	0.00113150
8	0.00061643	0.00035405	0.00025200	0.00026238	0.00036443
9	0.00013538	0.00006161	0.00003932	0.00007377	0.00009606
10	0.00002676	0.00000925	0.00000527	0.00001751	0.00002149
11	0.00000479	0.00000120	0.00000061	0.00000358	0.00000418
12	0.00000078	0.00000014	0.00000006	0.00000064	0.00000072
13	0.00000012	0.00000001	0.00000001	0.00000010	0.00000011
14	0.00000002	0.00000000	0.00000000	0.00000001	0.00000002

From the examples 3.1 and 3.2, it is seen that the improved binomial approximation is more appropriate than the binomial approximation.

4. Conclusion

The result of this study is an improved binomial distribution with parameters n and $p = \frac{r}{N}$ for approximating the Pólya distribution with parameters N, n, r and c. The improvement of the approximation is more appropriate than the binomial approximation, that is, the improved binomial distribution can be used as an approximation of the Pólya distribution when N is sufficiently large.

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] K. Teerapabolarn, Binomial approximation to Pólya distribution, *International Journal of Pure and Applied Mathematics*, **78** (2012), 635–640.