

## AN IMPROVED NEGATIVE BINOMIAL TO APPROXIMATE THE PÓLYA DISTRIBUTION

K. Teerapabolarn

Department of Mathematics

Faculty of Science

Burapha University

Chonburi, 20131, THAILAND

**Abstract:** This paper gives an improved negative binomial distribution with parameters  $r$  and  $p$  to approximate the Pólya distribution with parameters  $N$ ,  $m$  and  $r$ , where  $p = 1 - q = \frac{N}{N+m}$ . The improved approximation is more accurate than the negative binomial approximation when  $N$  is sufficiently large.

**AMS Subject Classification:** 62E17, 60F05

**Key Words:** negative binomial approximation, negative binomial probability function, Pólya probability function

### 1. Introduction

A ball is drawn at random from an urn containing  $r$  blue balls and  $N - r$  red balls. If a blue ball is drawn, then replace it together with another blue ball into this urn, and similarly for a red ball. Do this way for  $m$  draws and let  $X$  be the number of blue balls drawn in the  $m$  drawings, then the distribution of  $X$  is a Pólya distribution with parameters  $N$ ,  $m$  and  $r$ , and its probability function is given by

$$pY_{N,m,r}(x) = \frac{\binom{r+x-1}{x} \binom{N-r+m-x-1}{m-x}}{\binom{N+m-1}{m}}, \quad x = 0, 1, \dots, m. \quad (1.1)$$

The mean and variance of  $X$  are  $E(X) = \frac{rm}{N}$  and  $Var(X) = \frac{rm(N+m)(N-r)}{N^2(N+1)}$ ,

respectively. If  $N \rightarrow \infty$  while  $\frac{m}{N}$  remains a constant, then  $\mathbf{py}_{N,m,r}(x) \rightarrow \mathbf{nb}_{r,p}(x) = \binom{r+x-1}{x} p^r q^x$  for every  $x \in \{0, \dots, m\}$ , where  $p = 1 - q = \frac{N}{N+m}$  [2]. Therefore, the negative binomial probability function can be used as an estimate of the Pólya probability function when  $N$  is large. In this case, Malingam and Teerapabolarn [2] gave a bound on  $|\mathbf{py}_{N,m,r}(x) - \mathbf{nb}_{r,p}(x)|$  for  $x \in \{0, \dots, m\}$ .

In this paper, we focus on giving an improved negative binomial probability function,  $\widehat{\mathbf{nb}}_{r,p}(x)$ , to approximate the Pólya probability function, and the accuracy of the approximation is measured in the form of  $|\mathbf{py}_{N,m,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x)|$  for  $x \in \{0, \dots, m\}$ . 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 is directly obtained from [1].

**Lemma 2.1.** *For  $x, N \in \mathbb{N}$  and  $0 < p < 1$ , 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  $x \in \{0, \dots, m\}$ ,  $p = \frac{N}{N+m}$  and  $r \leq \sqrt{2N} - 1$ . Then we have the following:*

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

and for large  $N$ ,

$$\widehat{\mathbf{nb}}_{r,p}(x) \approx \mathbf{py}_{N,m,r}(x), \tag{2.4}$$

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

*Proof.* Let  $\eta = \begin{cases} 0 & \text{if } x = 0, \\ x - 1 & \text{if } x = 1, \dots, m. \end{cases}$  Applying Lemma 2.1, it follows

that

$$\mathbf{py}_{N,m,r}(x) = \binom{r+x-1}{x} \frac{[(N-1) \cdots (N-r)] [m \cdots (m-\eta)]}{(N+m-1) \cdots (N+m-r-x)}$$

$$\begin{aligned}
 &= \binom{r+x-1}{x} \frac{\prod_{i=1}^r \left(p - \frac{i}{N+m}\right) \prod_{i=0}^{\eta} \left(q - \frac{i}{N+m}\right)}{\prod_{i=1}^{r+x} \left(1 - \frac{i}{N+m}\right)} \\
 &= \binom{r+x-1}{x} \frac{p^r q^x}{1 + \frac{x(x-1)}{2(N+m)q}} \left\{ 1 + \frac{(r+x+1)(r+x)}{2(N+m)} \right. \\
 &\quad \left. - \frac{(r+1)r}{2(N+m)p} \right\} + O\left(\frac{1}{(N+m)^2}\right) \\
 &= \frac{\mathbf{nb}_{r,p}(x)}{1 + \frac{x(x-1)}{2m}} \left\{ 1 + \frac{(r+x+1)(r+x)}{2(N+m)} - \frac{(r+1)r}{2N} \right\} \\
 &\quad + O\left(\frac{1}{(N+m)^2}\right) \\
 &= \widehat{\mathbf{nb}}_{r,p}(x) + O\left(\frac{1}{(N+m)^2}\right).
 \end{aligned}$$

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

### 3. Numerical Examples

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

**3.1.** Let  $N = 50$ ,  $m = 10$  and  $r = 5$ , then  $p = \frac{50}{60}$  and the numerical results are as follows:

$x$	$\mathbf{py}_{N,m,r}(x)$	$\widehat{\mathbf{nb}}_{r,p}(x)$	$\mathbf{nb}_{r,p}(x)$	$ \mathbf{py}_{N,m,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x) $	$ \mathbf{py}_{N,m,r}(x) - \mathbf{nb}_{r,p}(x) $
0	0.38089033	0.38178369	0.40187757	0.00089337	0.02098725
1	0.35267623	0.35164288	0.33489798	0.00103335	0.01777825
2	0.17966525	0.17759741	0.16744899	0.00206784	0.01221626
3	0.06449522	0.06511905	0.06511905	0.00062383	0.00062383
4	0.01770457	0.01967138	0.02170635	0.00196681	0.00400178
5	0.00382419	0.00526379	0.00651191	0.00143960	0.00268772
6	0.00065037	0.00130238	0.00180886	0.00065201	0.00115849
7	0.00008517	0.00030564	0.00047375	0.00022048	0.00038858
8	0.00000815	0.00006909	0.00011844	0.00006093	0.00011028
9	0.00000051	0.00001519	0.00002851	0.00001467	0.00002800
10	0.00000002	0.00000327	0.00000665	0.00000325	0.00000664

**3.2.** Let  $N = 100$ ,  $m = 30$  and  $r = 10$ , then  $p = \frac{100}{130}$  and the numerical results are as follows:

$x$	$\mathbf{py}_{N,m,r}(x)$	$\widehat{\mathbf{nb}}_{r,p}(x)$	$\mathbf{nb}_{r,p}(x)$	$ \mathbf{py}_{N,m,r}(x) - \widehat{\mathbf{nb}}_{r,p}(x) $	$ \mathbf{py}_{N,m,r}(x) - \mathbf{nb}_{r,p}(x) $
0	0.06335387	0.06333139	0.07253815	0.00002249	0.00918428
1	0.15971565	0.16031360	0.16739573	0.00059796	0.00768008
2	0.21588683	0.21589065	0.21246381	0.00000382	0.00342301
3	0.20666089	0.20503501	0.19612044	0.00162589	0.01054045
4	0.15633184	0.15416198	0.14709033	0.00216986	0.00924151
5	0.09896485	0.09787600	0.09504298	0.00108886	0.00392187
6	0.05425705	0.05469189	0.05483249	0.00043485	0.00057544
7	0.02633970	0.02767935	0.02892263	0.00133965	0.00258293
8	0.01149422	0.01295112	0.01418321	0.00145689	0.00268899
9	0.00455627	0.00568778	0.00654610	0.00113151	0.00198983
10	0.00165268	0.00237124	0.00287021	0.00071855	0.00121753
11	0.00055135	0.00094654	0.00120428	0.00039518	0.00065293
12	0.00016975	0.00036417	0.00048635	0.00019443	0.00031660
13	0.00004832	0.00013575	0.00018993	0.00008743	0.00014161
14	0.00001273	0.00004923	0.00007201	0.00003650	0.00005928
15	0.00000310	0.00001743	0.00002659	0.00001433	0.00002348
16	0.00000070	0.00000604	0.00000959	0.00000534	0.00000889
17	0.00000015	0.00000205	0.00000338	0.00000191	0.00000324
18	0.00000003	0.00000069	0.00000117	0.00000066	0.00000114

From 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  $p = \frac{N}{N+m}$  was obtained. It is more appropriate for approximating the Pólya distribution, that is, the improved negative binomial distribution can be used as an estimate of the Pólya distribution when  $N$  is sufficiently large. In addition, the improved 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] P. Malingam, K. Teerapabolarn, A pointwise negative binomial approximation by  $w$ -functions, *International Journal of Pure and Applied Mathematics*, **69** (2011), 453–467.