An Improved Binomial Approximation for the Hypergeometric Distribution

K. Jaioun and K. Teerapabolarn*

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

* Corresponding author

Abstract

This paper gives an improved binomial distribution with parameters $m$ and $p$ to approximate the hypergeometric distribution with parameters $N$, $m$ and $n$, where $p = \frac{m}{N}$ and $m < n$. It is more accurate than the binomial approximation when $N$ is large.

Mathematics Subject Classification: 62E17, 60F05

Keywords: binomial approximation, binomial probability function, hypergeometric probability function

1 Introduction

Drawing $n$ balls at random, one at a time, without replacement from a box containing $N$ balls, $m$ ($m < n$) of which are white and $N - m$ are black. Let $X$ represents the number of white balls in the sample, then $X$ has a hypergeometric distribution. Its probability function can be expressed as
The mean and variance of $X$ are $E(X) = \frac{nm}{N}$ and $Var(X) = \frac{nm(N-m)(N-n)}{N^2(N-1)}$, respectively. It is well-known that if $N \to \infty$ and $p = \frac{m}{N}$ remains a constant then $h(x; m, n, N) \to \binom{n}{x} p^x (1-p)^{n-x}$ for $x \in \{0,1,\ldots,m\}$. Hence, the hypergeometric distribution can be approximated by a binomial distribution with parameters $n$ and $p = \frac{m}{N}$ [1]. In addition, it follows from [1] that if $N \to \infty$ and $p = \frac{n}{N}$ remains a constant then $h(x; m, n, N) \to b(x; m, p) = \binom{m}{x} p^x (1-p)^{m-x}$ for $x \in \{0,1,\ldots,m\}$. Thus, the hypergeometric distribution can be approximated by a binomial distribution with parameters $m$ and $p = \frac{n}{N}$.

In this paper, we focus on determining an improved binomial probability function to approximate the hypergeometric probability function. The accuracy of the approximation is measured in terms of the distance $|h(x; m, n, N) - \hat{h}(x; m, p)|$ for $x \in \{0,1,\ldots,m\}$. The desired result of this study is in Section 2. In Section 3, some numerical examples have been given to illustrate the improved binomial approximation, and the conclusion of this study is presented in the last section.

2 Results

Using the same idea detailed as in [2], the following lemma is also obtained.

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

$$
\prod_{i=0}^{x-1} \left( q - \frac{i}{N} \right) = q^x - \frac{x(x-1)}{2N} q^{x-1} + O\left( \frac{1}{N^2} \right) \quad (2.1)
$$
Improved binomial approximation

\[
\prod_{i=0}^{x-1} \left(1 - \frac{i}{Np}\right) = \frac{1}{1 + \frac{x(x-1)}{2Np} + O\left(\frac{1}{N^2}\right)} \quad (2.2)
\]

\[
\prod_{i=0}^{x-1} \left(1 - \frac{i}{N}\right) = 1 + \frac{x(x-1)}{2N} + O\left(\frac{1}{N^2}\right). \quad (2.3)
\]

**Theorem 2.1.** For \(x \in \{0, 1, \ldots, m\}\), if \(m \leq \sqrt{\frac{2N(N-n)}{n}}\) and \(p = 1 - q = \frac{n}{N}\), then we have the following:

\[
h(x; m, n, N) = \hat{b}(x; m, p) + O\left(\frac{1}{N^2}\right) \quad (2.4)
\]

and for the large \(N\),

\[
h(x; m, n, N) \approx \hat{b}(x; m, p), \quad (2.5)
\]

where \(\hat{b}(x; m, p) = b(x; m, p)\left\{1 + \frac{m(m-1)}{2N} - \frac{x(m-x)(m-x-1)}{2(N-n)}\right\} \bigg/ \left\{1 + \frac{x(x-1)}{2n}\right\} \).

**Proof.** Applying Lemma 2.1, it follows that

\[
h(x; m, n, N) = \begin{pmatrix} m \\ x \end{pmatrix} \frac{n!}{(n-x)!} \frac{(N-n)!}{(N-n-m+x)!} \frac{(N-m)!}{N!} \]

\[
= \begin{pmatrix} m \\ x \end{pmatrix} \left[\frac{n \cdots (n-(x+1))}{N(N-1)\cdots(N-m+1)}\right] \left[\frac{(N-n)\cdots(N-n-m+x+1)}{N}\right] \]

\[
= \begin{pmatrix} m \\ x \end{pmatrix} \left[\frac{n \cdots (n-x-1)}{N(N-1)\cdots(1-m-1)}\right] \left[\frac{1 \cdots (m-x-1)}{N(N-1)\cdots(1-m-1)}\right] \]

\[
= \begin{pmatrix} m \\ x \end{pmatrix} \prod_{i=0}^{x-1} \left\{p - \frac{i}{N}\right\} \prod_{i=0}^{m-1} \left\{q - \frac{i}{N}\right\} \prod_{i=0}^{m-1} \left\{1 - \frac{i}{N}\right\},
\]

where \(n \cdots (n-x+1) = \prod_{i=0}^{x-1} \left\{\frac{n-x-1}{N}\right\} = \prod_{i=0}^{x-1} \left\{p - \frac{i}{N}\right\} = 1 \) when \(x = 0\). By using Lemma 2.1, we have
K. Jaioun and K. Teerapabolarn

\[ h(x;m,n,N) = \frac{\binom{m}{x} p^x q^{m-x}}{1 + \frac{x(x-1)}{2Np} + O\left(\frac{1}{N^2}\right)} \left(1 - \frac{(m-x)(m-x-1)}{2Nq} + O\left(\frac{1}{N^2}\right)\right) \]
\[ \times \left(1 + \frac{m(m-1)}{2N} + O\left(\frac{1}{N^2}\right)\right) \]
\[ = \frac{b(x;m,p)}{1 + \frac{x(x-1)}{2n}} \left(1 + \frac{m(m-1)}{2N} - \frac{(m-x)(m-x-1)}{2(N-n)}\right) + O\left(\frac{1}{N^2}\right) \]
\[ = \hat{b}(x;m,p) + O\left(\frac{1}{N^2}\right). \]

Also, if \( N \) is large, then \( O\left(\frac{1}{N^2}\right) \approx 0 \). So \( h(x;m,n,N) \approx \hat{b}(x;m,p) \).

### 3 Numerical examples

The following numerical examples have been given to illustrate how well the improved binomial distribution approximates the negative hypergeometric distribution.

3.1 Let \( N = 100, n = 30, m = 10, p = \frac{n}{N} = 0.3 \) and the numerical results are as follows:

| \( x \) | \( h(x;m,n,N) \) | \( \hat{b}(x;m,p) \) | \( b(x;m,p) \) | \( | h(x;m,n,N) - b(x;m,p) | \) | \( | h(x;m,n,N) - \hat{b}(x;m,p) | \) |
|---|---|---|---|---|---|
| 0 | 0.02291724 | 0.02279979 | 0.02824752 | 0.00533028 | 0.00011745 |
| 1 | 0.11270774 | 0.11327834 | 0.12106082 | 0.00835308 | 0.00057060 |
| 2 | 0.23723161 | 0.23724016 | 0.23347444 | 0.00375717 | 0.00000854 |
| 3 | 0.28116339 | 0.27895647 | 0.26682793 | 0.01433546 | 0.00220692 |
| 4 | 0.20757766 | 0.20607693 | 0.20012095 | 0.00745671 | 0.00150073 |
| 5 | 0.09963728 | 0.10089772 | 0.10291935 | 0.00328207 | 0.00126044 |
| 6 | 0.03145116 | 0.03343128 | 0.03675691 | 0.00530575 | 0.00019802 |
| 7 | 0.00643776 | 0.00745098 | 0.00900169 | 0.00256393 | 0.00010132 |
| 8 | 0.00081655 | 0.00107434 | 0.00144670 | 0.00063015 | 0.00025778 |
| 9 | 0.00005876 | 0.00009081 | 0.00013778 | 0.00007993 | 0.00003295 |
| 10 | 0.00000174 | 0.00000342 | 0.00000590 | 0.00000417 | 0.00000169 |
3.2 Let \( N = 50, n = 10, m = 5, \) \( p = \frac{n}{N} = 0.5 \) and the numerical results are as follows:

<table>
<thead>
<tr>
<th>( x )</th>
<th>( h(x; m, n, N) )</th>
<th>( \tilde{b}(x; m, p) )</th>
<th>( b(x; m, p) )</th>
<th>( |h(x; m, n, N) - b(x; m, p)| )</th>
<th>( |\tilde{b}(x; m, n, N) - \tilde{b}(x; m, p)| )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.31056278</td>
<td>0.31129600</td>
<td>0.32768000</td>
<td>0.01711722</td>
<td>0.00073322</td>
</tr>
<tr>
<td>1</td>
<td>0.43133720</td>
<td>0.43008000</td>
<td>0.40960000</td>
<td>0.02173720</td>
<td>0.00125720</td>
</tr>
<tr>
<td>2</td>
<td>0.20983972</td>
<td>0.20945455</td>
<td>0.20480000</td>
<td>0.00503972</td>
<td>0.00038517</td>
</tr>
<tr>
<td>3</td>
<td>0.04417678</td>
<td>0.04627692</td>
<td>0.05120000</td>
<td>0.00702322</td>
<td>0.00210014</td>
</tr>
<tr>
<td>4</td>
<td>0.00396458</td>
<td>0.00480000</td>
<td>0.00640000</td>
<td>0.00243542</td>
<td>0.00083542</td>
</tr>
<tr>
<td>5</td>
<td>0.00011894</td>
<td>0.00019200</td>
<td>0.00032000</td>
<td>0.00020106</td>
<td>0.00007306</td>
</tr>
</tbody>
</table>

The numerical results in examples 3.1 and 3.2 indicate that the improved binomial approximation is more accurate than the binomial approximation.

4 Conclusion

In this study, an improved binomial distribution with parameters \( m \) and \( p = \frac{n}{N} \) for approximating the hypergeometric distribution with parameters \( N, n \) and \( m \) was obtained. It can be seen that the improved binomial distribution gives a good approximation for the hypergeometric distribution and also gives more accurate than the binomial distribution when \( N \) is large.

References


Received: December 18, 2013