On the Convergence of Hypergeometric to Binomial Distributions

Upul Rupassara[1], Bishnu Sedai[1]

[1] Department of Mathematics and Computer Science, Minot State University, Minot, USA

Correspondence: Upul Rupassara, Department of Mathematics and Computer Science, Minot State University, Minot, ND 58707, USA. E-mail: upul.rupassara@minotstateu.edu

Received: May 7, 2023 Accepted: June 20, 2023 Online Published: July 24, 2023

doi:10.5539/cis.v16n3p15 URL: https://doi.org/10.5539/cis.v16n3p15

Abstract

This study presents a measure-theoretic approach to estimate the upper bound on the total variation of the difference between hypergeometric and binomial distributions using the Kullback-Leibler information divergence. The binomial distribution can be used to find the probabilities associated with the binomial experiments. But if the sample size is large relative to the population size, the experiment may not be binomial, and a binomial distribution is not a good choice to find the probabilities associated with the experiment. The hypergeometric probability distribution is the appropriate probability model to be used when the sample size is large compared to the population size. An upper bound for the total variation in the distance between the hypergeometric and binomial distributions is derived using only the sample and population sizes. This upper bound is used to demonstrate how the hypergeometric distribution uniformly converges to the binomial distribution when the population size increases relative to the sample size.

Keywords: Hypergeometric, Binomial, Total variation, Upper bound, Kullback-Leibler

1. Introduction

Suppose that there are N elements in the population, each of which has one of two qualities: r successes and $N-r$ failures. If a random sample drawn from that population consists of n elements, y of which are successes, the hypergeometric probability distribution is given by

$$h_y^n(N,r) = \frac{\binom{r}{y}\binom{N-r}{n-y}}{\binom{N}{n}}, \tag{1}$$

where $y \in \mathbb{Z}_{\geq 0}$, subject to the restrictions $y \leq r$ and $r - y \leq N - n$. The expected value and variance of the hypergeometric distribution are respectively given by

$$E(Y) = \frac{nr}{N}, \tag{2}$$

and

$$V(Y) = n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right). \tag{3}$$

If the probability of success is p, the binomial distribution with parameters n and p is given by

$$b_y(n,p) = \binom{n}{y} p^y (1-p)^{n-y}, \tag{4}$$

where $y \leq n$, and $\binom{n}{y} = 0$ if $y > n$.

The expected value and variance of the binomial distribution are given by

$$E(Y) = np, \tag{5}$$

and

$$V(y) = np(1-p). \tag{6}$$

respectively.

The independence of the observation deteriorates and the experiment departs from being a binomial experiment when the sample size is large relative to the population. Because of this, the binomial distribution cannot provide a precise answer for the probabilities. The probability of success in the hypergeometric distribution is r/N, and it is a well-known result

that the hypergeometric distribution converges to the binomial distribution when the ratio $r/N = p$ stays constant as N increases.

$$\lim_{N\to\infty} h_y^n(N, r) = b_y(n, p). \tag{7}$$

Further, when the ratio r/N satisfies the above condition as N increases, by taking the limits in (2) and (3),

$$\lim_{N\to\infty} \frac{nr}{N} = np, \tag{8}$$

$$\lim_{N\to\infty} n\left(\frac{r}{N}\right)\left(\frac{N-r}{N}\right)\left(\frac{N-n}{N-1}\right) = np(1-p), \tag{9}$$

converge to the expected value and variance of the binomial distribution.

Kullback-Leibler divergence (Kullback and Leibler, 1951) is one of the most frequently used concepts when examining information divergence across various domains. If (X, S, μ_i) is a probability space for $i = 1, 2$, according to the Raydon-Nikodym theorem, for all $E \in S$, there exists $f_i(x), i = 1, 2$ such that

$$\mu_i(E) = \int_E f_i(x) d\lambda(x), \tag{10}$$

where λ is a probability measure. If H_1 and H_2 represent the hypothesis that the observation x is selected from the population 1 and 2 respectively, the mean information for discrimination $I(1 : 2, E)$ between H_1 and H_2 per observation from $E \in S$ for μ_1 is given by

$$I(1 : 2, E) = \frac{1}{\mu_1(E)} \int_E f_1(x) \log \frac{f_1(x)}{f_2(x)} d\lambda(x), \tag{11}$$

where $\mu_1(E) > 0$ and $I(1 : 2, E) = 0$ when $\mu_1(E) = 0$.

Literature Review

The hypergeometric distribution is required when sampling from a finite population without replacement. In this scenario, since trials are not independent and the probability of success of each trial (in general) is not constant, the binomial experiment is inapplicable because those conditions are required for a binomial experiment (Wroughton and Cole, 2013). This article further demonstrates how the probability mass functions behave differently depending on whether the experiment is performed with replacement or without replacement. When the experiment is carried out with replacements, the expected values of both distributions are the same as long as the ratio r/N remains constant. The variances are identical when $n = 1$, and they differ by a factor of $(N - n)/(N - 1)$ for $n > 1$. This difference in variances between two distributions reduces as the difference between n and N decreases. Because $n \leq N$, this discrepancy shrinks as n increases. As a result, the variance of the hypergeometric distribution decreases until it reaches $np(1 - p)$, which equals the variance of the binomial distribution.

In the recent literature, a pointwise upper bound for the distance between hypergeometric and binomial distributions has been derived in (Teerapabolan and Wongkasem, 2011). If $p = r/N$ and $q = 1 - p$, it is given by

$$|h_y^n(N, r) - b_y(n, p)| = \begin{cases} \frac{(n-1)(1-q^n)q}{N-1}, & \text{if } y = 0, \\ \min\left\{\frac{1-p^n}{y}, \frac{1-p^{n+1}-q^{n+1}}{(n+1)p}\right\}\frac{(n-1)np}{N-1}, & \text{if } 1 \leq y \leq n. \end{cases} \tag{12}$$

The improved binomial approximation is used to provide another improvement to this convergence problem in (Teerapabolan, 2014).

$$h_y^n(N, r) = \hat{b}_y(n, p) + O(1/N^2). \tag{13}$$

where $\hat{b}_y(n, p)$ is the modified binomial distribution given by

$$\hat{b}_y(n, p) = b_y(n, p)\left\{1 + \frac{n(n-1)}{2N} - \frac{(n-y)(n-y-1)}{2(N-r)}\right\}\Big/\left\{1 + \frac{y(y-1)}{2r}\right\}. \tag{14}$$

In addition to the upper bound for the distance, several scholars have investigated various methodologies and algorithms for computing probabilities for discrete distributions. Although we did not utilize that method to calculate the corresponding probability values, it is important noting for others who may be interested in expanding on this work in the future.

A recursive approach and algorithm with associated Fortran functions were discussed in (Berry et al, 1994). In this recursive technique, discrete distributions are formed as recursively defined positive functions, and the combination of recursion and a small random initial value ensures computing efficiency. According to their approach the cumulative probability p of x marked items is given by

$$P(x|n,p) = \sum_{j=0}^{x} r(j|n,p) \Big/ \sum_{j=0}^{n} r(j|n,p), \tag{15}$$

where $r(x|n,p)$ is a recursively defined positive function:

$$r(x+1|n,p) = r(x|n,p).\frac{(n-x)p}{(x+1)(1-p)}. \tag{16}$$

Since the hypergeometric distribution does not belong to the exponential family, many researchers have studied the information divergence of this distribution. Additionally, since this distribution is not exponential, approximating it from a binomial or Poisson distribution advances the field of information theory. The upper and lower bounds on the information divergence presented in (Harremoes and Matus, 2020) have some intriguing findings.

Our study provides a new perspective on prior studies on the topic by using a measure-theoretic approach to estimate an upper bound on the total variation of the difference between hypergeometric and binomial distributions. The result from (Teerapabolan and Wongkasem, 2011) gives precise values for the distances between the binomial and hypergeometric distributions, which depend on the number of variables used to establish those two distributions. The proposed method here starts with an upper bound for the total variation of the difference and uses it to compute the distance between two distributions. But we used a different approach using the information divergence concept and derived an upper bound for the total variation, which is a stronger result than the upper bound for just the distance between two distributions.

2. Method and Results

The primary objective of this article is to analyze the nature of the convergence and find an upper bound for the total variation of the difference between the hypergeometric and binomial distributions using the Kullback-Leibler mean information criterion. It is convenient to represent the hypergeometric probability distribution as in (17). Simplifying the (1) for $y \leq n$, we have

$$h_y^n(N,r) = \binom{n}{y}\left(\frac{r}{N}\right)^y \frac{\prod_{j=0}^{y-1}(1-j/r)\prod_{j=0}^{n-y-1}(1-p-j/N)}{\prod_{j=0}^{n-1}(1-j/N)}. \tag{17}$$

Furthermore, given the assumption that $r/N = p$, the following well-known result can be obtained from (17).

$$\lim_{N\to\infty} h_y^n(N,r) = \binom{n}{y} p^y (1-p)^{n-y}. \tag{18}$$

In order to get an upper bound for the total variation for the difference between the hypergeometric and binomial distributions, we use the general inequality derived using the total variation of the signed measure constructed based on the equivalence criteria of the distribution functions.

Let F_1 and F_2 be two distributions with density functions f_1 and f_2, respectively. Let v be a sigma finite signed measure of $F_1 - F_2$ that is absolutely continuous with respect to a sigma finite measure μ. According to Raydon-Nycodym theorem (Halmos, 1974),

$$v = \int (f_1 - f_2)d\mu. \tag{19}$$

further, the total variation of the signed measure v satisfies

$$\|v\| = \int |f_1 - f_2|d\mu. \tag{20}$$

With the use of Kullback-Leibler mean information criterion and inequality derived in (Ikeda, 1963),

$$\int |f_1 - f_2|d\mu \leq 2\left(\int f_1 \ln(f_1/f_2)d\mu\right)^{1/2}, \tag{21}$$

which proved to be a very useful inequality in (Vervaat, 1969) in terms of obtaining an upper bound for the total variation of the difference between binomial and Poisson distributions. When used for discrete distributions, μ is considered a counting measure, and integral is replaced by summation.

Proposition 1. *If* $0 < p < 1$ *and* $p = r/N$*, we have*

$$\frac{h_y^n(N,r)}{b_y(n,p)} \leq \frac{1}{\prod_{j=0}^{n-1}(1-j/N)}. \tag{22}$$

Proof. Using (4), (17), and simplifying, we get

$$\frac{h_y^n(N,r)}{b_y(n,p)} = \frac{\prod_{j=0}^{y-1}(1-j/r)\prod_{j=0}^{n-y-1}(1-p-j/N)}{(1-p)^{n-y}\prod_{j=0}^{n-1}(1-j/N)}. \tag{23}$$

The result in (22) followed from the facts,

$$\prod_{j=0}^{y-1}(1-j/r) \leq 1, \tag{24}$$

and

$$\prod_{j=0}^{n-y-1}(1-p-j/N) \leq (1-p)^{n-y}. \tag{25}$$

□

Also, according to the exposition given in (Hu et al, 2013), we have

$$\frac{1}{\prod_{j=0}^{n-1}(1-j/N)} = 1 + \frac{n(n-1)}{2N} + O(1/N^2). \tag{26}$$

Proposition 2. *If* $0 < p < 1$ *and* $p = r/N$*, we have*

$$\frac{h_y^n(N,r)}{b_y(n,p)} \leq 1 + \frac{n(n-1)}{2N} + O(1/N^2). \tag{27}$$

Proof. The result is directly followed from Proposition 1 and (26). □

Further, as a direct consequence of Propositions 1 and 2, we have

$$\lim_{N\to\infty} \frac{h_y^n(N,r)}{b_y(n,p)} = 1. \tag{28}$$

The Propositions 1 and 2 set an upper bound on the ratio $h_y^n(N,r)/b_y(n,p)$, which is only dependent on sample and population sizes. However, there might be a sharper upper bound that can be described in terms of all or a few other variables used to establish each distribution. Figures 1, 2, and 3 demonstrate how the hypergeometric distribution behaves as sample sizes vary and population sizes increase. In all three scenarios, p and y are both set to constants. In each figure, the horizontal dashed line represents the binomial probability for the parameters n and p. In Figure 1, $b_{10}(20, 0.3) = 0.0308$, while $b_{10}(30, 0.3) = 0.142$, and $b_{10}(40, 0.3) = 0.113$ in Figures 2 and 3 respectively. The hypergeometric distribution asymptotically converges to the binomial probability in each case as N increases, while the behavior is convoluted for small N.

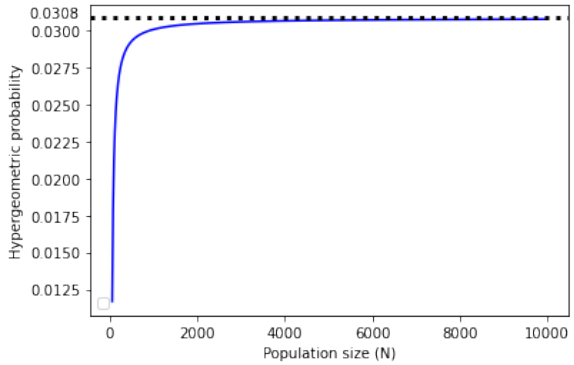

Figure 1. Hypergeometric probability vs population size for $p = 0.3$, $y = 10$, $n = 20$.

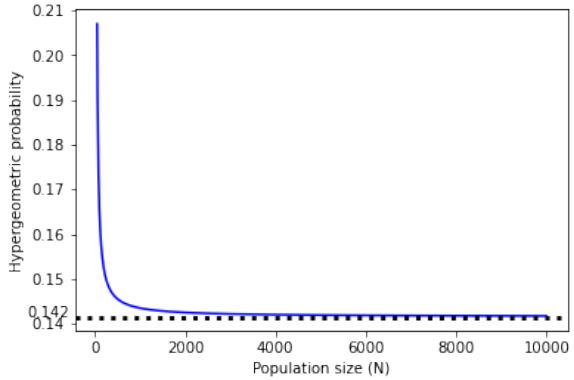

Figure 2. Hypergeometric probability vs population size for $p = 0.3$, $y = 10$, $n = 30$.

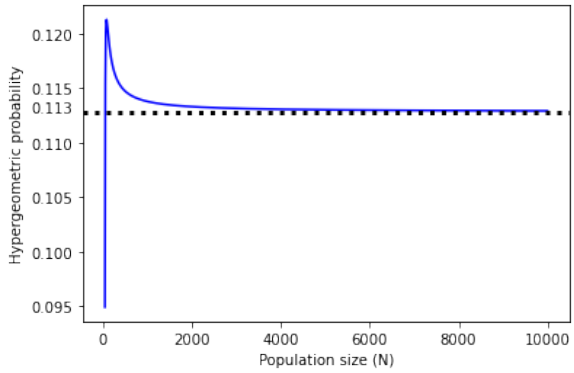

Figure 3. Hypergeometric probability vs population size for $p = 0.3$, $y = 10$, $n = 40$.

Theorem 1. *Let $h_y^n(N, r)$ and $b_y(n, p)$ be hypergeometric and binomial distributions respectively. Suppose the ratio r/N remains as a constant (p) when N approaches infinity and $0 < p < 1$. Then we have*

$$\sum_{y=0}^{n} |h_y^n(N, r) - b_y(n, p)| \leq \frac{2(n-1)}{\sqrt{N-(n-1)}} \tag{29}$$

Proof. Using the discrete version of (21), we get

$$\sum_{y=0}^{n} |h_y^n(N, r) - b_y(n, p)| \leq 2\left(\sum_{y=0}^{n} h_y^n(N, r) \ln \frac{h_y^n(N, r)}{b_y(n, p)} \right)^{1/2}. \tag{30}$$

By the result derived in Proposition 1,

$$\sum_{y=0}^{n} h_y^n(N,r)\ln\frac{h_y^n(N,r)}{b_y(n,p)} \leq \sum_{y=0}^{n} h_y^n(N,r)\sum_{j=0}^{n-1}\ln\Big(\frac{1}{1-j/N}\Big). \tag{31}$$

Using the logarithmic inequalities shown in (Love, 1980), and by the fact

$$\sum_{y=0}^{n} h_y^n(N,r) = 1, \tag{32}$$

the inequality (30) can be written as

$$\begin{aligned}\sum_{y=0}^{n}|h_y^n(N,r)-b_y(n,p)| &\leq 2\sqrt{\sum_{j=0}^{n-1}\Big(\frac{j}{N-j}\Big)} \\ &\leq \frac{2(n-1)}{\sqrt{N-(n-1)}}.\end{aligned} \tag{33}$$

□

Corollary 1. *When n is sufficiently small compared to N, we have*

$$\lim_{N\to\infty}\sum_{y=0}^{n}|h_y^n(N,r)-b_y(n,p)| = 0 \tag{34}$$

Proof. The result follows from Theorem 1. □

4. Discussion

In this article, we proposed a technique to use the Kullback-Leibler information criterion to establish an upper bound on the total variation for the distance between hypergeometric and binomial distributions. We considered the case where the support of the hypergeometric distribution is $y = 0, 1, ..., n$. However, if $n \geq r$, then the hypergeometric distribution is defined for $y = 0, 1, ..., r$. When the population size N increases, r increases as well. But the ratio r/N remains a constant value. We only considered small sample sizes in comparison to the population size. The distance between various probability distributions as one converges to another has been analyzed using a variety of methodologies that have been developed by many researchers. We used similar techniques used in (Vervaat, 1969). As a result of finite sampling in the binomial distribution in our situation, the total variance also comprises a finite summation, making it possible to derive the upper bound without the use of intricate mathematical results and principles.

5. Conclusion

A number of scholars have explored the convergence scenario of hypergeometric to binomial distributions using various methodologies. In this article, we presented a novel technique based on a measure-theoretic approach and the idea of Kullback-Leibler information divergence. This method is used to derive a uniform upper bound for the total variation of the distance between those two distributions. Even though the total variation entails a finite summation, this method enables one to determine an upper bound for the distance between two distributions that entirely depends on the population and sample sizes. In contrast to previous findings and methodologies found in the recent literature, the techniques used in this work and the results derived, make a significant contribution to other related investigations over comparable studies.

References

Berry, K. J., Mielke, P. W., & Helmericks, S. G. (1994). An algorithm to generate discrete probability distributions: Binomial, hypergeometric, negative binomial, inverse hypergeometric, and Poisson. *Behavior Research Methods, Instruments, and Computers*, *26,* 366-367. https://doi.org/10.3758/BF03204645.

Halmos, P. R. (1974). Measure theory / [by] Paul R. Halmos. Springer-Verlag New York. Retrieved from http://www.loc.gov/catdir/enhancements/fy0814/74010690-t.html

Harremo?s, P., & Mats, F. (2020). Bounds on the Information Divergence for Hypergeometric Distributions. *Kybernetika*, *56,* 1111-1132. https://doi.org/10.14736/kyb-2020-6-1111

Hu, D. P., Cui, Y. Q., & Yin, A. H. (2013). An Improved Negative Binomial Approximation for Negative Hypergeometric Distribution. *Applied Mechanics and Materials*, *427-429,* 2549-2553. https://doi.org/10.4028/www.scientific.net/amm.427-429.2549

Ikeda, S. (1963). Asymptotic equivalence of probability distributions with applications to some problems of asymptotic independence. *Annals of the Institute of Statistical Mathematics*, *15,* 87-116, http://doi.org/10.1007/BF02865908

Kullback, S., & Leibler, R. A. (1951). On Information and Sufficiency. *The Annals of Mathematical Statistics*, *22*(1), 79-86. https://doi.org/10.1214/aoms/1177729694

Love, E. R. (1980). Some Logarithm Inequalities. *The Mathematical Gazette*, *64*(427), 55-57. https://doi.org/10.2307/3615890

Teerapabolarn, K. (2014). An improved binomial to approximate the hypergeometric distribution. *International journal of pure and applied mathematics*, *90,* 515-518. https://doi.org/10.12732/ijpam.v91i1.8

Teerapabolarn, K., & Wongkasem, P. (2011). On pointwise binomial approximation by *w*-functions. *International journal of pure and applied mathematics*, *71,* 57.

Vervaat (1969). Upper bounds for the distance in total variation between the binomial or negative binomial and the poisson distribution. *Statistica Neerlandica*, *23*(1), 79-86. https://doi.org/10.1111/j.1467-9574.1969.tb00075.x

Wroughton, J., & Cole, T. (2013). Distinguishing Between Binomial, Hypergeometric and Negative Binomial Distributions. *Journal of Statistics Education*, *21*(1). https://doi.org/10.1080/10691898.2013.11889663