Pure Mathematics 理论数学, 2022, 12(9), 1419-1428
Published Online September 2022 in Hans. http://www.hanspub.org/journal/pm
https://doi.org/10.12677/pm.2022.129155

Hans汉斯

二项分布的正态近似若干条件

孙舒婷

云南财经大学，云南 昆明

收稿日期：2022年7月31日；录用日期：2022年8月31日；发布日期：2022年9月7日

摘 要

二项分布是统计问题中常见的分布，对于试验次数n较大时，直接计算的困难很大，因此常用泊松分布或正态分布近似进行计算。本文主要研究二项分布的正态近似，依据中心极限定理，当试验次数n足够大时，可以把服从二项分布的独立同分布的随机变量之和当作正态变量，从而利用正态分布对二项分布近似计算。理论上，二项分布的正态近似只要求n充分大即可，但实际应用中，在n相对较大时，参数p及随机变量取值k的不同对近似计算准确性的影响较为明显。本文用实验法，通过对比独立重复的二项分布与正态近似概率的相对误差得知，对于不同参数p，n值充分大的程度不同。为了更好地得到近似的结果，应根据p的取值，确定n的最小下限值，同时，使用正态近似连续性修正公式可以进一步提高计算的准确性。相较于其他学者的研究，针对不同参数条件的二项分布，本文给出了确定n的最小下限值的方法，一定程度上减少了资源的浪费，为不同试验中n的“充分大”划出了较为明确的定义。

关键词

二项分布，中心极限定理，正态近似

Conditions for Normal Approximation of Binomial Distribution

Shuting Sun

Yunnan University of Finance and Economics, Kunming Yunnan

Received: Jul. 31st, 2022; accepted: Aug. 31st, 2022; published: Sep. 7th, 2022

Abstract

Binomial distribution is a common distribution in statistical problems. When the number of tests is large, it is very difficult to calculate directly. Therefore, Poisson distribution or normal distribution approximation is often used for calculation. This paper mainly studies the normal approxi-

文章引用: 孙舒婷. 二项分布的正态近似若干条件[J]. 理论数学, 2022, 12(9): 1419-1428.
DOI: 10.12677/pm.2022.129155

mation of the binomial distribution. According to the central limit theorem, when the number of tests is large enough, the sum of the random variables that obey the binomial distribution and are independent and identically distributed can be regarded as the normal variable, so that the binomial distribution can be approximated by the normal distribution. Theoretically, the normal approximation of binomial distribution can only be obtained if *n* is sufficiently large. However, in practical application, when *n* is relatively large, the difference in the value of parameter *p* and random variable *k* has obvious influence on the accuracy of approximate calculation. By comparing the relative error between the binomial distribution of independent repetition and the normal approximation probability, we can know that the minimum lower limit value of *n* should be determined according to the value of *p* for different parameters with different degrees of sufficient magnitude, so as to obtain better approximation results. At the same time, the accuracy of calculation can be further improved by using the continuity correction formula of normal approximation. Compared with the research of other scholars, this paper gives a method to determine the minimum lower limit of *n* for the binomial distribution under different parameter conditions, which reduces the waste of resources to a certain extent, and draws a more clear definition for the "sufficiently large" of *n* in different experiments.

Keywords

Binomial Distribution, Central Limit Theorem, Normal Approximation

1. 引言

1.1. 研究背景及意义

二项分布是统计问题中常见的离散型随机变量的分布，在实际生活中，很多现象都可以用二项分布进行描述，例如抛硬币得到正反面问题、彩票是否中奖、生男生女问题等。

事实上，在 n 次独立重复的伯努利(Bernoulli)试验中，设每次试验中事件 A 发生的概率为 p。用 X 表示 n 重伯努利试验中事件 A 发生的次数，则 X 的可能取值为 $0,1,\cdots,n$，且对每一个 $k\left(0\le k\le n\right)$，事件 $\left\{X=k\right\}$ 即为"n 次试验中事件 A 恰好发生 k 次"，随机变量 X 的离散概率分布即为二项分布(Binomial distribution)。

此时 n 次试验中事件 A 恰好发生 k 次的概率由概率质量函数(Probability mass function)给出(盛骤等，1989) [1]，如下：

$$P\left\{X=k\right\}=\binom{k}{n}p^k\left(1-p\right)^{n-k}$$

根据上述公式可以看出，当 n 取值较小时，二项分布的计算难度一般，然而实例中，经常利用多次试验避免偶然性，因此 n 的取值通常较大，直接计算的工作量十分大，计算难度较高，需要使用简便高效的计算方法进行计算。

为了解决上述问题，国内外学者通常对二项分布采用基于泊松分布的近似方法以及基于中心极限定理的近似方法。本文主要基于中心极限定理研究二项分布的正态近似问题与正态近似中的若干条件。

正态分布(Normal distribution)，也称"常态分布"，又名高斯分布(Gaussian distribution)，最早由棣

莫弗(Abraham de Moivre)在求二项分布的渐近公式中得到，是一个在数学、物理及工程等领域都非常重要的概率分布，在统计学的许多方面有着重大的影响力。

若随机变量 X 服从一个数学期望为 μ 、方差为 σ^2 的正态分布，记为 $N\left(\mu,\sigma^2\right)$。其概率密度函数为正态分布的期望值 μ 决定了其位置，其标准差 σ 决定了分布的幅度。

对于一维正态分布，若随机变量 X 服从一个位置参数为 μ 、尺度参数为 σ 的概率分布，且其概率密度函数为 $f\left(x\right)=\dfrac{1}{\sqrt{2\pi}\sigma}\exp\left(\dfrac{\left(x-\mu\right)^2}{2\sigma^2}\right)$，则这个随机变量就称为正态随机变量，正态随机变量服从的分布就称为正态分布，记作 $X\sim N\left(\mu,\sigma^2\right)$，读作 X 服从 $N\left(\mu,\sigma^2\right)$，或 X 服从正态分布。

n 维随机向量具有类似的概率规律时，称此随机向量遵从多维正态分布。多元正态分布有很好的性质，例如，多元正态分布的边缘分布仍为正态分布，它经任何线性变换得到的随机向量仍为多维正态分布，特别它的线性组合为一元正态分布。当 $\mu=0$， $\sigma^2=1$时，正态分布就成为标准正态分布。

基于正态分布良好的性质，许多学者在研究统计模型时通常采用正态近似进行探究。

1.2. 国内外研究

1.2.1. 用泊松分布近似二项分布

二项分布作为统计问题中常见的离散型随机变量的分布，应用十分广泛，但是由于 n 的取值通常较大，计算难度较高，需要更为高效的计算方法。在一定条件下，依托泊松定理，可以用泊松分布近似二项分布。

在马小霞(2007)的研究中[2]，当参数间的关系为 $\lambda\approx np$ 时，二项分布的极限分布是泊松分布。这里要求 n 充分大时，np 不大或 $n(1-p)$ 不大，也就是 p 比较小，才能得到比较精确的近似值。

在侯国亮(2018)的研究中[3]曾提到，在实践中，一般当 $n\geq 20$， $p\leq 0.05$ 时，用 $\dfrac{\lambda^k\mathrm{e}^{-\lambda}}{k!}$ 作为 $C_n^k p^k\left(1-p\right)^{n-k}$ 的近似值效果较好。

吴艳华(2012)曾依据试验数据说明[4]，当 n 很大 p 很小且 $np\left(1-p\right)\leq 9$ 时，用泊松近似比正态近似效果要好，并且当 $n\geq 100$， $p\leq 0.01$时，二项分布的泊松近似精度更高。这一结论在侯国亮(2018) [3]针对具体问题的研究中也得到了印证。在李斐等(2020) [5]的研究中，也对两种近似的收敛速度进行了比较，可以根据不同的参数条件，选择较为适合的近似方法。

相对于泊松近似对 n 较大、p 较小的要求，正态近似只要求 n 充分大，因此二项分布的正态近似普适性更强，且正态分布性质较好，有许多学者也针对其进行了更深层次的研究。

1.2.2. 用中心极限定理近似二项分布

对于独立同分布的随机变量，中心极限定理如下：

设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 独立同分布，并且具有有限的数学期望和方差：$E\left(X_i\right)=\mu$，$D\left(X_i\right)=\sigma^2$ $\left(i=1,2,\cdots\right)$，则对任意 x，分布函数 $F_n\left(x\right)=P\left\{\dfrac{\sum_{i=1}^n X_i-n\mu}{\sigma\sqrt{n}}\leq x\right\}$，满足

$$\lim_{n\to\infty}F_n\left(x\right)=\lim_{n\to\infty}\left\{\frac{\sum_{i=1}^n X_i-n\mu}{\sqrt{n}\sigma}\leq x\right\}=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}\mathrm{e}^{-\frac{t^2}{2}}\mathrm{d}t=\Phi\left(x\right)$$

该定理说明，当 n 很大时，随机变量 $Y_n=\dfrac{\sum_{i=1}^n X_i-n\mu}{\sqrt{n}\sigma}$ 近似地服从标准正态分布 $N\left(0,1\right)$。因此，当 n

很大时，$\sum_{i=1}^{n} X_i = \sqrt{n}\sigma Y_n + n\mu$ 近似地服从正态分布 $N\left(n\mu, n\sigma^2\right)$。

该定理是中心极限定理最简单又最常用的一种形式，在实际工作中，只要 n 足够大，便可以把独立同分布的随机变量之和当作正态变量。这种方法在数理统计中用得很普遍，当处理大样本时，它是重要工具。

如果 n 足够大，那么分布的偏度就比较小。在这种情况下，如果使用适当的连续性校正，那么 $B(n,p)$ 的一个很好的近似是正态分布：$N\left(np, np(1-p)\right)$。

不同的经验法则可以用来决定 n 是否足够大，以及 p 是否距离 0 或 1 足够远，盛骤等(1989)描述为[4]：当 n 充分大时，可作近似计算。在《概率统计》(2000) [6]一书中则定义为：正态分布作近似计算，它的优点是不受"$p \le 0.1$"限制，只要 n 足够大即可。

其中一个常用的规则是 np 和 $n(1-p)$ 都必须大于 5。当 n 越大(至少 20)且 p 不接近 0 或 1 时近似效果更好。

Chang *et al*. (2006, 2008)表明[7] [8]，应用中心极限定理时，样本大小应随概率分布类型而变化。

2. 研究内容及研究方法

为了探究不同的 n 与 p 对近似效果的影响，本文利用统计模拟衡量不同条件下近似值与精确值之间的差距。

对独立重复的二项分布实验，每次发生概率为 p 的事件 A 的总次数 $X_i\left(i=1,2,\cdots,n\right)$，若 n 次独立重复实验中事件 A 发生 k 次，此时 $P\left(X \le k\right) = \sum_0^k C_n^k p^k \left(1-p\right)^{n-k}$，以此可以求得二项分布下的精确概率值 $P_{精确}$。

对于上述实验，本文中按照正态近似公式求得此时 $P\left(X \le k\right)$，如下：

$$\lim_{n\to\infty} P\left\{\frac{X-np}{\sqrt{np\left(1-p\right)}} \le x\right\} = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-\frac{t^2}{2}} \mathrm{d}t = \Phi\left(x\right)，\text{即 } P_{近似}。$$

这里可以利用相对误差衡量 $P_{精确}$ 与 $P_{近似}$ 之间的差距，如下：

$$相对误差 = \left|P_{精确} - P_{近似}\right|$$

$$相对误差(\%) = \frac{\left|P_{精确} - P_{近似}\right|}{P_{精确}} \times 100$$

3. 统计模拟

对 $n = 10 \sim 300$ 范围内步长为 10 的每一 n 值相应的不同参数 $p = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5$ 时的概率 $P\left(X \le k\right)$ 进行了大量的对比计算，文中仅给出部分计算结果，见表 1，同理可以做概率 $P\left(X = k\right)$ 时的对比计算。

Table 1. Relative error and corresponding k value of normal approximation $P\left(X = k\right)$ with different parameters n and p

表 1. 不同参数 n，p 正态近似 $P\left(X = k\right)$ 的相对误差及相应 k 值

$n = 50, p = 0.01$				$n = 50, p = 0.1$			
k	精确值	近似值	相对误差(%)	k	精确值	近似值	相对误差(%)
0	0.605006067	0	100	0	0.005153775	0	100
1	0.910564687	0.52271066	42.59489004	1	0.03378586	0.020462157	39.43573775
2	0.986182729	0.744852201	24.4711777	2	0.111728756	0.069438541	37.85078876
3	0.998403827	0.761165147	23.76179589	3	0.250293906	0.16367823	34.60558709

Continued

4	0.999854311	0.761355003	23.85340595	4	0.431198407	0.309464881	28.23144138
5	0.999989103	0.76135533	23.86363738	5	0.616123008	0.490788937	20.34237788
6	0.999999315	0.76135533	23.86441484	6	0.770226842	0.672112993	12.73830556
7	0.999999963	0.76135533	23.8644642	7	0.877854916	0.817899644	6.829747272
8	0.999999998	0.76135533	23.86446688	8	0.942132794	0.912139334	3.183570375
9	1	0.76135533	23.86446701	9	0.975462064	0.961115718	1.470723152
10	1	0.76135533	23.86446701	10	0.990645398	0.981577875	0.915314792
11	1	0.76135533	23.86446701	11	0.996780079	0.98845007	0.83569177
12	1	0.76135533	23.86446701	12	0.99899538	0.990305223	0.869889617
13	1	0.76135533	23.86446701	13	0.99971488	0.990707717	0.90097317
14	1	0.76135533	23.86446701	14	0.999926161	0.990777892	0.914894484
15	1	0.76135533	23.86446701	15	0.999982503	0.990787723	0.919494092
16	1	0.76135533	23.86446701	16	0.999996197	0.99078883	0.920740274
17	1	0.76135533	23.86446701	17	0.99999924	0.99078893	0.921031784
18	1	0.76135533	23.86446701	18	0.99999986	0.990788937	0.921092477
	$n = 50, p = 0.5$				$n = 300, p = 0.01$		
k	精确值	近似值	相对误差(%)	k	精确值	近似值	相对误差(%)
0	8.88E−16	0	100	0	0.049040894	0	100
1	4.53E−14	4.91E−12	10733.67801	1	0.197649664	0.082057668	58.48327455
2	1.13E−12	3.80E−11	3251.164138	2	0.422063918	0.240008659	43.13452318
3	1.85E−11	2.44E−10	1214.970751	3	0.647233775	0.459138624	29.06139299
4	2.23E−10	1.43E−09	539.6450308	4	0.816111167	0.678268589	16.89017189
5	2.10E−09	7.71E−09	266.1820025	5	0.917096437	0.836219579	8.818795289
6	1.62E−08	3.85E−08	137.3876346	6	0.967249054	0.918277248	5.062998585
7	1.05E−07	1.78E−07	69.56462718	7	0.988525921	0.948996158	3.998859615
8	5.82E−07	7.61E−07	30.8052282	8	0.996397288	0.9572805	3.925822366
9	2.81E−06	3.01E−06	7.332600314	9	0.998976905	0.958889366	4.01285941
10	1.19E−05	1.10E−05	7.421363735	10	0.999735156	0.959114277	4.063164012
11	4.51E−05	3.75E−05	16.85055103	11	0.999937078	0.959136899	4.08027463
12	1.53E−04	0.000118017	22.83027301	12	0.999986199	0.959138536	4.084822704
13	0.000468111	0.000344257	26.45833731	13	0.999997191	0.959138621	4.085868503
14	0.001301086	0.000931423	28.4118542	14	0.999999467	0.959138624	4.086086499
15	0.003300224	0.002338867	29.13003779	15	0.999999905	0.959138624	4.086128535
16	0.007673339	0.005454749	28.91296421	16	0.999999984	0.959138624	4.0861361

Continued

17	0.016419569	0.011825808	27.97735139	17	0.999999998	0.959138624	4.086137377
18	0.032454324	0.02385744	26.48917766	18	1	0.959138624	4.08613758
19	0.059460226	0.044843011	24.58318158	19	1	0.959138624	4.08613758
20	0.101319376	0.078649604	22.37456744	20	1	0.959138624	4.08613758
21	0.16111816	0.128949518	19.96587008	21	1	0.959138624	4.08613758
22	0.239943831	0.198071955	17.45069924	22	1	0.959138624	4.08613758
23	0.335905517	0.285803822	14.91541278	23	1	0.959138624	4.08613758
24	0.443862414	0.388648705	12.43937458	24	1	0.959138624	4.08613758
25	0.556137586	0.5	10.09419031	25	1	0.959138624	4.08613758
26	0.664094483	0.611351295	7.942121167	26	1	0.959138624	4.08613758
27	0.760056169	0.714196178	6.033763489	27	1	0.959138624	4.08613758
28	0.83888184	0.801928045	4.40512509	28	1	0.959138624	4.08613758
29	0.898680624	0.871050482	3.074522958	29	1	0.959138624	4.08613758
30	0.940539774	0.921350396	2.040251543	30	1	0.959138624	4.08613758
31	0.967545676	0.955156989	1.280424031	31	1	0.959138624	4.08613758
32	0.983580431	0.97614256	0.756203672	32	1	0.959138624	4.08613758
33	0.992326661	0.988174192	0.418457911	33	1	0.959138624	4.08613758
34	0.996699776	0.994545251	0.216165916	34	1	0.959138624	4.08613758
35	0.998698914	0.997661133	0.103913377	35	1	0.959138624	4.08613758
36	0.999531889	0.999068577	0.046352868	36	1	0.959138624	4.08613758
37	0.999847068	0.999655743	0.019135421	37	1	0.959138624	4.08613758
38	0.999954893	0.999881983	0.007291305	38	1	0.959138624	4.08613758
39	0.999988069	0.999962493	0.002557624	39	1	0.959138624	4.08613758
40	0.999997193	0.999988955	0.000823822	40	1	0.959138624	4.08613758
41	0.999999418	0.999996987	0.00024311	41	1	0.959138624	4.08613758
42	0.999999895	0.999999239	6.56E−05	42	1	0.959138624	4.08613758
43	0.999999984	0.999999822	1.62E−05	43	1	0.959138624	4.08613758
44	0.999999998	0.999999961	3.64E−06	44	1	0.959138624	4.08613758
45	1	0.999999992	7.49E−07	45	1	0.959138624	4.08613758
46	1	0.999999999	1.41E−07	46	1	0.959138624	4.08613758
47	1	1	2.44E−08	47	1	0.959138624	4.08613758
48	1	1	3.95E−09	48	1	0.959138624	4.08613758
49	1	1	6.44E−10	49	1	0.959138624	4.08613758
50	1	1	1.54E−10	50	1	0.959138624	4.08613758

通过计算分析得出：

1) n 值相同，p 值不同时，近似计算的准确性并不相同。在 $p \le 0.5$ 达到范围内，p 越接近 0.5，近似计算的准确性越高，即 $p = 0.5$ 时的近似效果最好。由二项分布的特点可知，$p \ge 0.5$ 范围内，近似计算的准确性变化与上述相同。二项分布只在 $p = 0.5$ 时为对称分布，此时，用正态近似准确性最高[9]。

2) 对于 $P(X = k)$ 接近于 0 时，k 值在随机变量的均值 np 附近相对误差较小，近似计算的准确性较高。但需注意 $P(X = k)$ 接近于 0 时，单纯通过相对误差的数值不能确切说明近似程度，还应看 $P(X = k)$ 本身概率值的大小。

3) 对于不同参数 n、p 时，正态近似 $P(X = k)$ 的最大绝对误差(近似值与精确值之差的绝对值)及相应的 k 值，见表 2，仅给出部分结果；当允许误差不超过 0.004 时，不同参数 p 正态近似所需 n 的最小下限值见表 3。

可以发现，当 $p < 0.1$ 或 $p > 0.9$ 时，要达到好的近似效果，正态近似所需 n 很大，此时应考虑其他方法进行近似。

Table 2. Absolute error and corresponding k value of normal approximation $P(X = k)$ with different parameters n and p

表 2. 不同参数 n，p 正态近似 $P(X = k)$ 的绝对误差及相应 k 值

p	$n = 30$		$n = 50$	
	最大误差	相应 k	最大误差	相应 k
0.01	0.1106	0	0.1645	0
0.05	0.0629	0	0.0413	1
0.1	0.0259	2	0.018	3
0.2	0.0125	4	0.0071	8
0.3	0.0061	7	0.0036	12
0.4	0.0027	9	0.0016	17
0.5	0.0012	15	0.0006	25
p	$n = 90$		$n = 210$	
	最大误差	相应 k	最大误差	相应 k
0.01	0.1365	0	0.0502	1
0.05	0.021	3	0.009	8
0.1	0.0095	7	0.004	18
0.2	0.004	15	0.0017	37
0.3	0.0019	23	0.0008	58
0.4	0.0009	39	0.0004	88
0.5	0.0003	45	0.0001	105

Continued

p	n = 300	
	最大误差	相应 k
0.01	0.0306	1
0.05	0.0062	12
0.1	0.0028	26
0.2	0.0012	55
0.3	0.0006	96
0.4	0.0003	126
0.5	0.0005	150

Table 3. The minimum lower limit of n required for normal approximation of different parameters p
表 3. 不同参数 p 正态近似所需 n 的最小下限值

不同参数 p 正态近似所需 n 的最小下限值					
p	$0.1 \le p < 0.2$	$0.2 \le p < 0.3$	$0.3 \le p < 0.4$	$0.4 \le p < 0.5$	
n	210	90	50	30	
p	$0.5 \le p < 0.6$	$0.6 \le p < 0.7$	$0.7 \le p < 0.8$	$0.8 \le p < 0.9$	$0.9 \le p < 1$
n	30	30	50	90	210

例如图 1，当 $n = 50$， $p = 0.1$ 时，将精确值用线(红色)连接起来可以看到，其形状和近似值(蓝色)较为相近，并且在 $k = 3$ 处的绝对误差最大，并逐渐趋向于 0。

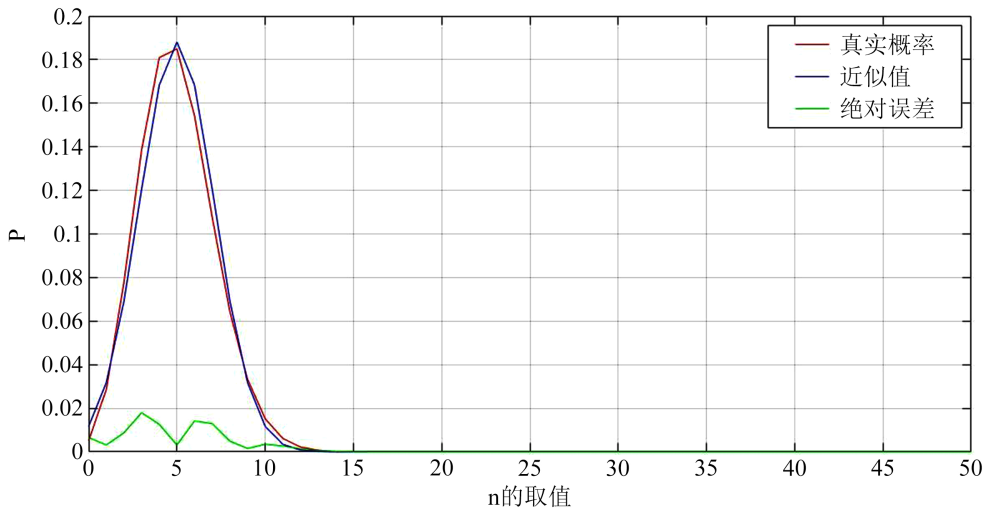

Figure 1. The exact value and approximate value of $P(X = k)$ when n = 50, p = 0.1 after normal approximation continuity, the exact value and approximate value of k
图 1. 正态近似连续性修正后 n = 100，p = 0.01 时，$P(X = k)$ 的精确值与近似值

例如图 1，当 $n = 50$，$p = 0.1$ 时，将精确值用线(红色)连接起来可以看到，其形状和近似值(蓝色)较为相近，并且在 $k = 3$ 处的绝对误差最大，并逐渐趋向于 0。

由上述图表可以看出，通过中心极限定理对二项分布进行正态近似的效果较为明显，但是误差仍然波动不定。这里对正态近似进行连续性修正，即 $P(X \le k) = P(X < k+1) = \Phi\left\{\frac{k+\frac{1}{2}-np}{\sqrt{np(1-p)}}\right\}$。同样可以利用 MATLAB 实现对修正后公式的误差验证。

这里只展示 $n = 100$，$p = 0.01$ 时的情况。

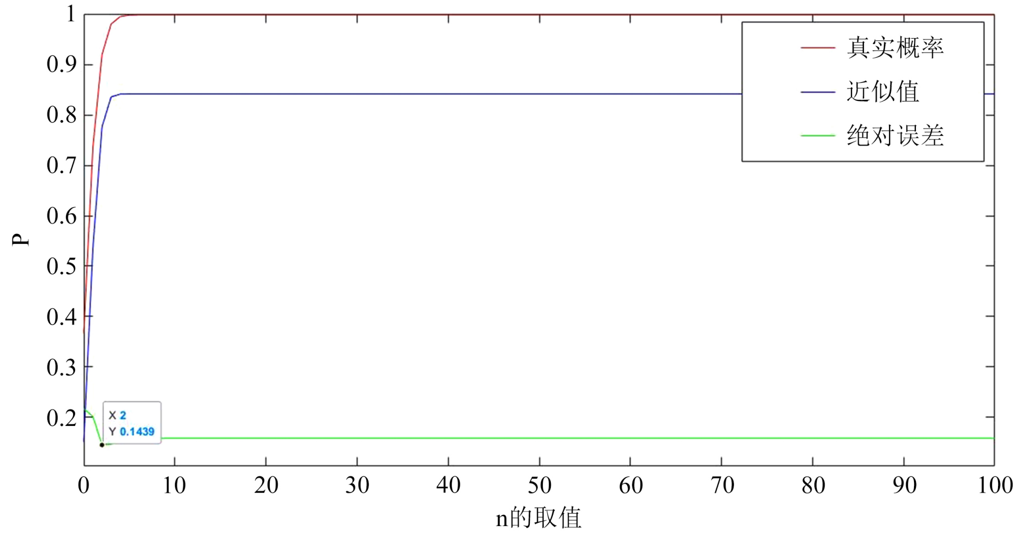

Figure 2. The exact value and approximate value of $P(X \le k)$ when $n = 100$, $p = 0.01$ after normal approximation continuity

图 2. 正态近似连续性修正后 $n = 100$，$p = 0.01$ 时，$P(X \le k)$ 的精确值与近似值

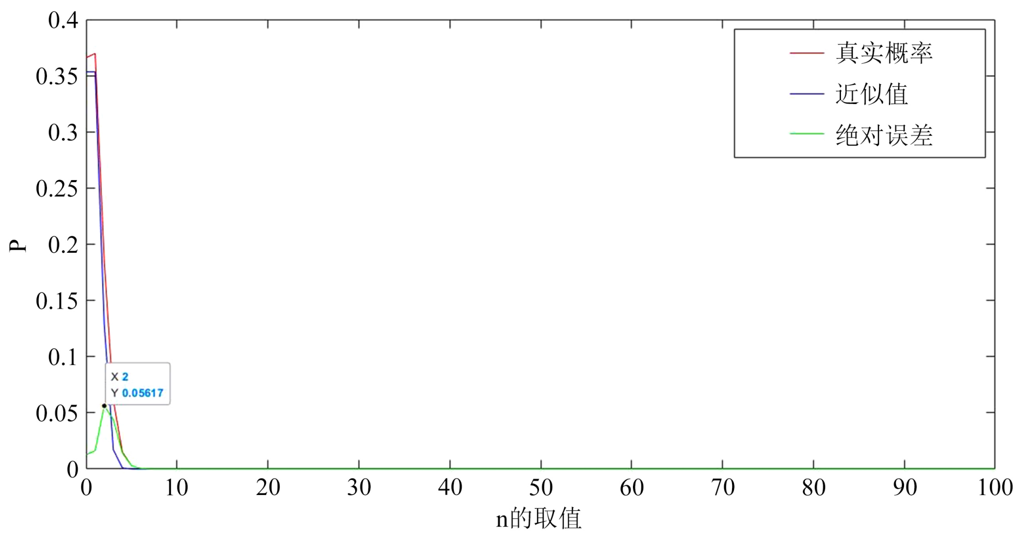

Figure 3. The exact value and approximate value of $P(X = k)$ when $n = 100$, $p = 0.01$ after normal approximation continuity

图 3. 正态近似连续性修正后 $n = 100$，$p = 0.01$ 时，$P(X = k)$ 的精确值与近似值

如图 2 所示，修正后，当 $n = 100$，$p = 0.01$ 时，对于概率 $P(X \le k)$ 在 $k = 2$ 时误差最小为 0.1439，此后基本趋于平缓。而图 3 中，在修正后，当 $n = 100$，$p = 0.01$ 时，对于概率 $P(X = k)$ 在 $k = 2$ 时误差最大为 0.05617，此后误差逐渐减小。可以看出，修正后的精确值与估计值较为吻合，近似效果较好。

4. 结论和讨论

4.1. 研究的结论

在过往的理论学习中，二项分布的正态近似只要求 n 充分大，而通过统计模拟后可以看出，在 n 相对较大时，参数 p 及随机变量取值 k 的不同对近似计算准确性的影响是比较明显的。

对于不同参数 p，n 值充分大的程度不同，为了更好地得到近似的结果，应用时可参考文中给出的不同 p 所需 n 的最小下限值。

当 $p < 0.1$ 或 $p > 0.9$ 时，要达到好的近似效果，正态近似所需 n 很大，此时应考虑其他方法进行近似。

二项分布使用正态近似时可以通过使用连续性修正近似公式，以此来提高近似的效果及计算的准确度。

4.2. 研究的局限性

本文有以下几个不足之处：1) 实际生活中，某一具体事件发生的概率并不一定是可知的，因此在概率 p 的确定上会有一些浮动，对最终 n 的取值也会有一定的影响，因此文中所给的 n 的最小下限值仅供参考。2) 文中的方法对随机变量 k 的不同取值及其对 n 的最小下限值影响的探究效果不明确，需要更好的方法进行持续的研究。

参考文献

[1] 盛骤, 谢式千, 潘承毅. 概率论与数理统计[M]. 北京: 高等教育出版社, 1989.

[2] 马小霞. 有关二项分布的近似计算[J]. 楚雄师范学院学报, 2007, 22(3): 23-26.

[3] 侯国亮. 二项分布事件概率的两种近似计算方法[J]. 现代职业教育, 2018(1): 182.

[4] 吴艳华. 谈概率论中二项分布的泊松近似与正态近似[J]. 城市建设理论研究: 电子版, 2012(33): 1-4.

[5] 李斐, 陈晓娜. 二项分布两种近似分布的收敛速度[J]. 高等数学研究, 2020, 23(4): 44-46.

[6] 同济大学概率统计教研组. 概率统计[M]. 上海: 同济大学出版社, 2000.

[7] Chang, H.J., Huang, K.C. and Wu, C.H. (2006) Determination of Sample Size in Using Central Limit Theorem for Weibull Distribution. *International Journal of Information and Management Sciences*, **17**, 31-46.

[8] Chang, H.J., Wu, C.H., Ho, J.F. and Chen, P.Y. (2008) On Sample Size in Using Central Limit Theorem for Gamma Distribution. *International Journal of Information and Management Sciences*, **19**, 153-174.

[9] 张子贤. 关于二项分布的正态近似计算问题[J]. 河北工程技术高等专科学校学报, 2002(1): 20-23.