

Assessing the Accuracy of the Normal Approximation to the Binomial Distribution in Large Samples: An Analytical Study and Monte Carlo Simulation

Alaa Farhan Ahmed Azawi
University of Tikrit, Iraq



DOI : <https://doi.org/10.61796/ipteks.v3i4.627>



Sections Info

Article history:

Submitted: June 07, 2026

Final Revised: July 16, 2026

Accepted: August 04, 2026

Published: September 20, 2026

Keywords:

Binomial distribution

Natural approximation

Large samples

Continuity correction

Monte Carlo simulation

Approximation accuracy

ABSTRACT

Objective: This study aims to evaluate the accuracy of the natural approximation of the binomial distribution in large samples through an analytical study and Monte Carlo simulations, focusing on the effect of sample size n and probability parameter P , as well as the continuity correction, and certain properties related to the shape of the distribution. **Method:** The study included 143 scenarios resulting from the use of sample sizes ranging from 10 to 10,000 and different values of P ranging from 0.01 to 0.99, with the calculation of a set of accuracy metrics, including absolute error, relative error, mean absolute error (MAE), root mean square error (RMSE), and maximum cumulative error, along with an examination of the relationship between the error and each of (nP) and $n(1-P)$, as well as the absolute skewness of the binomial distribution. Monte Carlo simulations with 10,000 iterations were also used to empirically verify the computational results. **Results:** It was also observed that there is a significant decrease in errors through continuity correction, where the mean absolute error reduced from (0.0362579686) without correction to (0.007330879) with the correction; the root mean square error also reduced from (0.0690117985) to (0.0162159958), which is an improvement in the maximum cumulative error of (79.539328%) on the set of correct values. **Novelty:** The study concludes that the assessment of the accuracy of the natural approximation should not depend solely on a large n , but should take into account P , (nP) , $n(1-P)$, and the shape parameters of the distribution, with the importance of using a continuity correction when applying the natural approximation to the binomial distribution.

INTRODUCTION

The binomial distribution is one of the most important discrete probability distributions in statistics, due to its fundamental role in describing the number of times a specific event occurs in a fixed number of independent trials, where each trial has two possible outcomes—success or failure—and the probability of success remains constant from one trial to the next. If x represents the number of successes in n independent trials, and the probability of success in a single trial is P , then x follows a binomial distribution of the form $x \sim \text{bin}(n, P)$, and the mean and variance be, respectively, $E(x) = nP$ and $\text{Var}(x) = nP(1-P)$. These properties have made the binomial distribution an important tool in a wide range of statistical applications that require estimating probabilities associated with the number of successes or events occurring within a given sample size [1,2]. Despite the simplicity of the binomial distribution's definition and its theoretical and practical importance, the direct calculation of its probabilities can become computationally more difficult as n increases, especially when dealing with a large number of scenarios or performing frequent comparisons. For this reason, the normal approximation to the

binomial distribution is of great importance, as it allows the discrete distribution with a continuous normal distribution with mean nP and variance $nP(1-P)$. This approximation is based on the possibility of representing the binary variable as a sum of independent Bernoulli variables, and thus taking advantage of the central limit theorem, which states that the standard normal distribution of this sum approaches the standard normal distribution as the sample size increases [3,4]. Consequently, if $x \sim \text{bin}(n, P)$, the normal approximation can be written as $x \sim N(nP, nP(1-P))$, or in terms of the standard variable $Z = \frac{x-nP}{\sqrt{nP(1-P)}}$. However, this approximation is a limit approximation and does not imply that every large value of n necessarily leads to a high-precision approximation in practical applications; the quality of the approximation can also be affected by the value of P , by the values nP and $n(1-P)$, by the shape of the binomial distribution and its symmetry, as well as by the location of the probability to be calculated within the distribution. Therefore, moving from the general statement that the binomial distribution approaches the normal distribution with large samples to an actual assessment of the extent of this approximation requires a precise quantitative study of the error resulting from the approximation [5]. The second difficulty that is caused by the fact that two types of distribution have distinct properties: the binomial distribution has a discrete property and consists of integers 0, 1, 2, ..., n , while the normal distribution has a continuous feature. Thus, the use of the normal distribution instead of the binomial one in order to calculate probabilities can produce discrepancies between the exact value and the approximate one [6]. In order to reduce the impact of such discrepancies a continuity correction should be performed, which includes the shifting of the boundaries of the event by 0.5 during the transition of probability from a discrete to a continuous distribution. Thus, the problem of the need to determine the degree of improvement that could be reached by means of a continuity correction appears, particularly the question of how this improvement depends on the sample size, probability parameter, and its position. The research problem is the necessity of a proper assessment of the quality of the natural approximation of the binomial distribution in large samples. The purpose of this research is to analyze the magnitude of the error in probabilities obtained with the use of normal approximation compared to those which were calculated exactly using the binomial distribution formula; to consider normal approximation with and without continuity correction; to explore the dependence of accuracy on parameters of n and P , as well as on their products nP and $n(1-P)$; and to analyze the relation between the accuracy of the approximation and the skewness of the binomial distribution. At the same time, the problem of distinguishing the error of the approximation from the random error in the simulation procedure is considered so that the analysis would be more clear and consistent. In order to do so, the study uses the numerical-analytical method, which consists of calculating the exact value, making the probabilistic approximations, and running the simulation procedure. The experiments involve varying sample size n from 10 to 10,000, and P -value from low to high, starting from zero till one. For each particular

case, the exact value of the probability was found using the binomial distribution and then the normal approximation was performed. The difference between the mentioned values is estimated based on the absolute error, relative error, mean absolute error, and root mean square error, in addition to the comparison of the cumulative distribution functions based on the maximum cumulative difference, with the help of other measures to investigate the influence of continuity correction. The research involves the application of the Monte Carlo simulation to confirm the estimated probabilities empirically through obtaining a large amount of random values from the binomial distribution and estimation of probabilities using the relative frequency of appearance of the events of interest. The Monte Carlo simulation is not used to replace the actual value of the binomial distribution; rather, it is used to confirm it independently. This approach enables a comparative analysis of three types of sources of results: an exact probability, a standard approximation as something that needs to be analyzed, and Monte Carlo simulation as a source of evidence for checking the approximation's validity. Not being restricted to just the analysis of sample size and the probability parameter, this study also analyzes such parameters as nP and $n(1-P)$, as well as skewness coefficient of the binomial distribution, which characterizes the asymmetry of the binomial distribution. This will enable us to identify whether those cases where n is large but P is close to zero or one differ in their approximation from those where P is close to 0.5. The comparison also enables us to identify whether the increased approximation accuracy due to larger sample sizes happens equally in all cases, or whether the improvement depends on the form of distribution. Therefore, the purpose of the research is to conduct a thorough quantitative analysis of the normal approximation accuracy to the binomial distribution, taking into account four dimensions: the influence of sample size, the influence of the probability value, the influence of the continuity correction, and the influence of the skewness value of the distribution. In other words, the purpose of the research is not to prove once again the theoretically known fact about the normal approximation accuracy of the binomial distribution for large values of n , but to find out the approximation accuracy under certain conditions.

RESEARCH METHOD

A numerical analysis technique, together with Monte Carlo simulation was applied in order to estimate the correctness of natural approximation of binomial distribution. It was assumed that variable x is distributed binomially, that is, $x \sim \text{bin}(n, P)$. The probabilities found through direct computation using binomial distribution were compared with the probabilities for natural approximation without and with continuity correction for sample sizes n varying from 10 to 10,000 with different values of P changing from 0.99 to 0.01. The correctness of the approximation was analyzed by means of absolute error, relative error, mean absolute error, root mean square error and maximal cumulative error, taking into account the influence of parameters nP , $n(1-P)$ and $|\gamma_1|$.

Additionally, Monte Carlo simulation was applied for empirical verification of the results with 10,000 iterations. All computations were made using the help of MATLAB.

The Theoretical Basis of the Binomial Distribution and the Normal Approximation

The Binomial Distribution

If x represents the number of successes in n independent trials, and the probability of success in each trial is constant and equal to P , then x follows the binomial distribution and is written as $x \sim \text{bin}(n, P)$. The probability mass function of the binomial distribution is given by the formula: [7]

$$P(X = x) = \binom{n}{x} P^x (1 - P)^{n-x} \quad x = 0, 1, 2, \dots, n \quad (1)$$

Where n represents the number of trials, P represents the probability of success in a single trial, $(1-P)$ represents the probability of failure, and x represents the number of successes in n trials. The binomial distribution is characterized by a mean and a variance given, respectively, by the following two relationships:

$$E(x) = nP \quad (2)$$

$$\text{Var}(x) = nP(1 - P) \quad (3)$$

Therefore, the standard deviation of a binomial distribution is given by the following formula:

$$\sigma = \sqrt{nP(1 - P)} \quad (4)$$

These characteristics form the basis for the normal approximation of the binomial distribution in large samples, since the normal approximation uses the same mean and variance.

The Normal Approximation of the Binomial Distribution

When the sample size is sufficiently large, the binomial distribution can be approximated using the normal distribution with the same mean and variance. Accordingly, if $x \sim \text{bin}(n, P)$, the normal approximation is written as:

$x \sim N(nP, nP(1-P))$ This approximation allows the discrete calculations of the binomial distribution to be replaced with calculations based on the cumulative distribution function of the normal distribution, which facilitates the calculation of various probabilities, especially for large values of n [8].

Theoretical Basis of the Approximation

The normal approximation to the binomial distribution can be explained by expressing the random variable x as the sum of n independent random variables following a Bernoulli distribution, such that $X = X_1 + X_2 + \dots + X_n$, where each X_i follows a Bernoulli distribution with a success probability of P for each of these variables [9]

$$E(X_i) = P$$

$$\text{Var}(X_i) = P(1 - P)$$

And thus it follows that

$$E(x) = nP$$

$$\text{Var}(x) = nP(1 - P)$$

According to the central limit theorem, the standard normal distribution of the sum of independent random variables approaches the normal distribution as n increases; therefore, the standard normal variable associated with the binomial distribution is written as:

$$Z = \frac{x - nP}{\sqrt{nP(1-P)}}$$

As n increases, the Z distribution approaches the standard normal distribution, which is the theoretical basis for using the normal approximation in the present study.

Continuity Correction

The binomial distribution differs from the normal distribution in that the former is discrete and the latter is continuous. Therefore, replacing the binomial distribution with the normal distribution may result in a difference between the exact probability and the approximate probability. To minimize this difference, a continuity correction is applied, which involves adjusting the probability limits by 0.5, for example, when calculating the cumulative probability: [10]

$$P(X \leq k)$$

The natural approximation is given without continuity correction:

$$P(X \leq k) \approx \Phi \left[\frac{(k - nP)}{\sqrt{nP(1-P)}} \right] \quad (5)$$

However, when using the continuity correction, it becomes:

$$P(X \leq k) \approx \Phi \left[\frac{(k + 0.5 - nP)}{\sqrt{nP(1-P)}} \right] \quad (6)$$

Conditions for the Suitability of the Natural Approximation of the Binomial Distribution

The natural approximation of the binomial distribution is considered more suitable when both the expected number of successes, nP , and the expected number of failures, $n(1-P)$, are sufficiently large, so that the binomial distribution is not severely skewed or excessively concentrated near either end of its range [11]. This is related to the properties of the binomial distribution: its mean is nP , while its variance is $nP(1-P)$; the skewness coefficient reflects the degree of asymmetry in its shape. For small values of P near 0.5, the distribution becomes more symmetrical around the mean, helping the normal approximation become better; however, for small values of P nearing zero or one, the approximation becomes difficult [12,13]. However, these conditions do not represent strict limits by which the accuracy of the approximation can be judged in all cases, because the magnitude of the error may vary depending on the sample size, the value of P , and the location of the probability within the distribution. Therefore, the present study does not rely on a single threshold condition to judge the validity of the approximation, but rather tests its accuracy quantitatively across a wide range of values of n and P , examining np and $n(1-P)$ as well as the absolute skewness, and comparing the exact results of the binomial distribution with the normal approximation with and without the continuity correction.

Measures for Evaluating the Accuracy of the Natural Approximation

To quantitatively evaluate the accuracy of the natural approximation of the binomial distribution, this study relies on a set of measures that compare the probabilities calculated directly from the binomial distribution—as reference values—with the probabilities resulting from the natural approximation, both with and without continuity correction. Absolute Error is used as the basic metric and it is computed using the equation below: [14]

$$AE = |P_{exact} - P_{approx}| \quad (7)$$

Where:

P_{exact} : The probability calculated directly from the exact binomial distribution.

P_{approx} : The probability obtained using the approximation method.

The relative error is also calculated as a supporting measure according to:

$$RE = \frac{|P_{exact} - P_{approx}|}{P_{exact}} \quad (8)$$

In combining the results of several instances, the mean absolute error (MAE) is used in order to determine the average value of the difference between the true values and the estimated values, whereas the root mean square error (RMSE) is used to emphasize larger errors:

$$MAE = \left(\frac{1}{m}\right) \sum |P_{(exact)i} - P_{(approx)i}| \quad (9)$$

$$RMSE = \sqrt{\left(\frac{1}{m}\right) \sum (P_{(exact)i} - P_{(approx)i})^2} \quad (10)$$

Where:

m : the number of observed cases

The accuracy of the approximation on the level of the cumulative distribution function as a whole is evaluated using the largest discrepancy between the cumulative distribution function of the binomial distribution and the cumulative distribution function of the normal approximation. The Kolmogorov measure for the uncorrected normal approximation is determined by the formula:

$$D_K = \sup_{x \in R} |F_B(x) - F_N(x)| \quad (11)$$

Where

$F_B(x)$: The cumulative distribution function for the binomial distribution.

$F_N(x)$: The Cumulative Distribution Function for the Approximate Normal Distribution.

The maximum error is also determined for the exact values of the distribution in case of approximation by the normal distribution with continuity correction and without continuity correction to be able to compare how well the approximation works in all cases. This combination of criteria makes it possible to estimate the quality of approximation locally and globally (cumulative distribution functions), and to see how much better the approximation works with continuity correction.

Monte Carlo Simulation and Its Role in Evaluating Approximations

For verification of the probabilities estimated from the binomial distribution by Monte Carlo simulation method as a numerical approach, this research makes use of Monte Carlo simulation technique. The basis of the Monte Carlo simulation is based on the generation of numerous random numbers for the binomial distribution for different combinations of the parameters n and P [15]. The generated random numbers are then employed for estimation of the probabilities and comparison of them with the actual values. For each scenario of this study, where x obeys the binomial distribution $\text{bin}(n, P)$, $M = 10,000$

independent observations are generated. To estimate the probability of the event $X \leq k$, the indicator function I is used, which takes the value 1 when the event occurs and 0 otherwise. The Monte Carlo probability estimate is therefore given by:

$$\hat{P}_{MC} = (1/M) \sum_{j=1}^M I(X_j \leq k) \quad (12)$$

Where:

X_j : the j th simulation observation.

M : the number of simulation runs.

As the number of runs increases, the simulated estimate becomes more stable on average.

Monte Carlo simulation is not used in this study as a substitute for exact probability, but rather to empirically verify the results. Therefore, the simulated value $\hat{P}_{MC}$ is compared with the exact value P_{exact} calculated directly from the binomial distribution. The simulation error is calculated as follows:

$$AE_{MC} = |\hat{P}_{MC}(X \leq k) - P_{\text{exact}}(X \leq k)| \quad (13)$$

The approximate standard error of the simulated capacity is calculated as follows:

$$SE_{MC} = \sqrt{\frac{\hat{P}_{MC}(X \leq k)[1 - \hat{P}_{MC}(X \leq k)]}{M}} \quad (14)$$

This allows us to distinguish between errors resulting from randomness in the simulation and those resulting from the use of natural approximation. Accordingly, the main comparison of approximation accuracy remains between P_{exact} and natural approximation, while the Monte Carlo results serve as a means of verifying the consistency of the computational results. A base simulation size of 10,000 iterations was used, with an additional stability check of the results using different simulation sizes in a limited number of scenarios: 1,000, 5,000, 10,000, 50,000, and 100,000 iterations. It will help to understand how the value of the estimated simulation changes with an increase in the number of iterations and whether the results that have been used for comparison are affected by the particular size of simulation used.

Experimental Design of the Numerical Study

The numerical study was designed to allow for an evaluation of the natural approximation of the binomial distribution across a wide range of sample sizes and probability parameters. The study included sample sizes

$n = 10, 20, 30, 50, 100, 200, 500, 1,000, 2,000, 5,000, 10,000$, while the P-values

$P = 0.01, 0.05, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 0.95, 0.99$. Such a set of values allows the creation of 143 experimental situations where it is possible to investigate the influence of both n and P on the accuracy of the approximation. For each situation of values of n and P , the mean value nP and the standard deviation $\sqrt{nP(1-P)}$ were computed at the first stage, and then the exact probability of the binomial distribution was calculated as a reference point. After that, the probability was calculated by the normal approximation without the continuity correction, and then by the normal approximation with the continuity correction. Points of evaluation corresponding to different positions relative to the mean of the distribution were selected based on the values of the standard scores $Z = -2, -1, 0, 1, 2$, omitting duplicate values caused by rounding. In addition to n and P , the values nP and $n(1-P)$ were calculated, and the coefficient of skewness of the binomial distribution was calculated by the formula:

$$\gamma_1 = \frac{(1-2P)}{\sqrt{nP(1-P)}} \quad (15)$$

The absolute value $|\gamma_1|$ is employed to examine the relationship between the form of distribution and the amount of the rounding error to explain the differences in the rounding precision in different cases. The design of the experiment produced the exact and approximate probabilities of each case, as well as the absolute and relative errors, results of the Monte Carlo simulation, MAE and RMSE of the overall error, the maximal cumulative error and the Kolmogorov distance of the uncorrected approximation. The improvement gained due to the continuity correction was found as well. All calculations and simulations were implemented with the help of MATLAB software with the use of a fixed-seed random number generator. The exact value of the binomial distribution was chosen as the reference in the study, while the Monte Carlo simulation was used to prove the results; however, simulations were not considered the replacement of the exact computation.

RESULTS AND DISCUSSION

Result

From the numerical analysis, it was noted that the accuracy of the natural approximation of the binomial distribution increased as the sample size was increased; however, the rate of increase depended on the probability parameter P . When $P = 0.5$, the maximum cumulative error of the natural approximation of the binomial distribution without continuity correction was found to drop from 0.12305 at $n = 10$ to 0.0039893 at $n = 10,000$, while the error after continuity correction dropped from 0.0026862 to 2.7152×10^{-5} .

6. It was observed that the accuracy of approximation is greater in the central region of P values, and the error increases as P approaches either zero or one, suggesting that increasing the sample size is not enough to determine the accuracy of approximation. From the comparison of the various scenarios considered, it was seen that the use of continuity correction led to reduced error metrics in comparison to natural approximation without continuity correction; the average absolute error and root mean square error were 0.03625797 and 0.0690118 without continuity correction and 0.007330879 and 0.016216, respectively. The mean maximum cumulative rounding error for the set of exact values without correction was 0.07937393, whereas the mean maximum cumulative rounding error with correction was 0.01695333, with a mean improvement of 79.5393%. The results suggest a marked improvement in the precision of the rounding procedure when applying the continuity correction in the given context. Additionally, based on the analysis of the skewness coefficient, there is a notable positive relationship between the absolute skewness value $|\gamma_1|$ and the rounding error; the Spearman correlation coefficient was 0.82912 in the case of rounding without continuity correction and 0.98189 for rounding with continuity correction. In turn, negative correlations have been found between the error and both nP and $n(1-P)$, the degree of which depends on the approximation method used. However, the Monte Carlo simulation indicated good correspondence between probabilities from the simulation and probabilities of the binomial distribution with a certain amount of randomness depending on the size of the simulation. Therefore, the simulation validated the computational part of the paper without substituting the exact probabilities, which were used as a standard of comparison. In general, one can say that in order to evaluate the natural approximation, it is necessary to consider not only the continuity correction but also sample size, P -value, values of nP and $n(1-P)$, and skewness. The table below provides values of MAE, RMSE, mean D_{Grid} and mean improvement ratio.

Table 1. Summary of accuracy measures for the normal approximation to the binomial distribution

The Scale	Without continuity correction	With the confirmation of continuity
<i>MAE</i>	0.03625797	0.007330879
<i>RMSE</i>	0.0690118	0.016216
Average D_{Grid}	0.07937393	0.01695333
Average improvement	-	% 79.5393

To visually illustrate the similarity between the binomial distribution and the normal distribution, the following figure compares the probability mass function of the binomial distribution with the probability density function of the normal distribution when $n = 100$ and $P = 0.5$.

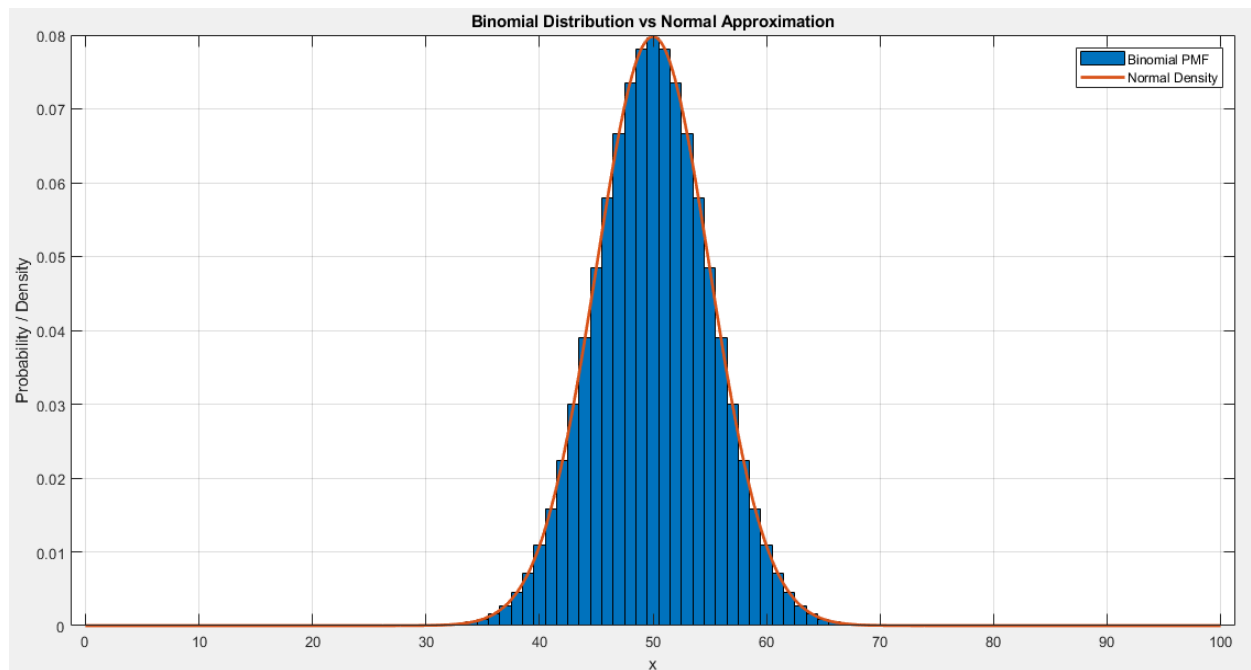


Figure 1. Comparison of the binomial distribution with the normal approximation at $n = 100$ and $P=0.5$

The following figure shows a comparison of the cumulative distribution functions for the binomial distribution and the two normal approximations.

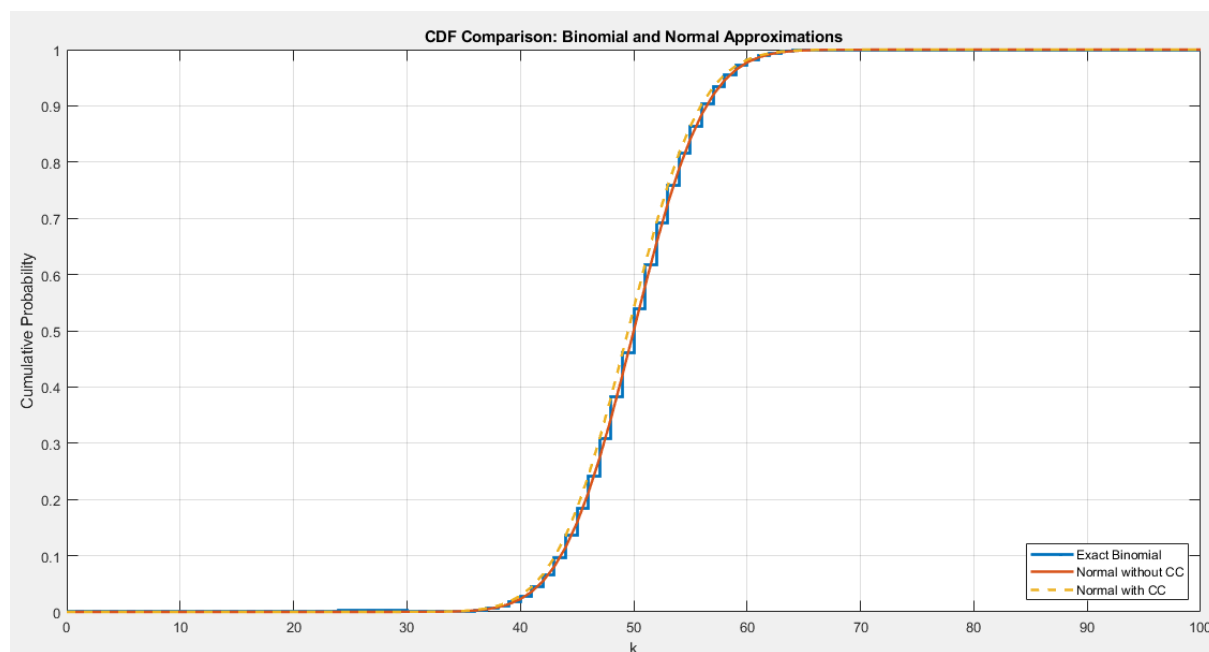


Figure 2. Shows a comparison of the cumulative distribution functions for the binomial distribution and the two normal approximations.

The results showed that increasing the sample size leads to a decrease in the approximation error. This result is clearly evident when $P = 0.5$ is fixed: for $n = 10$, $(D_{Grid,CC})$ ranged from 0.0026862 to 2.7152×10^{-6} . These results indicate that the

approximation accuracy improves as n increases. The following figure illustrates the effect of sample size n on approximation accuracy when $P = 0.5$.

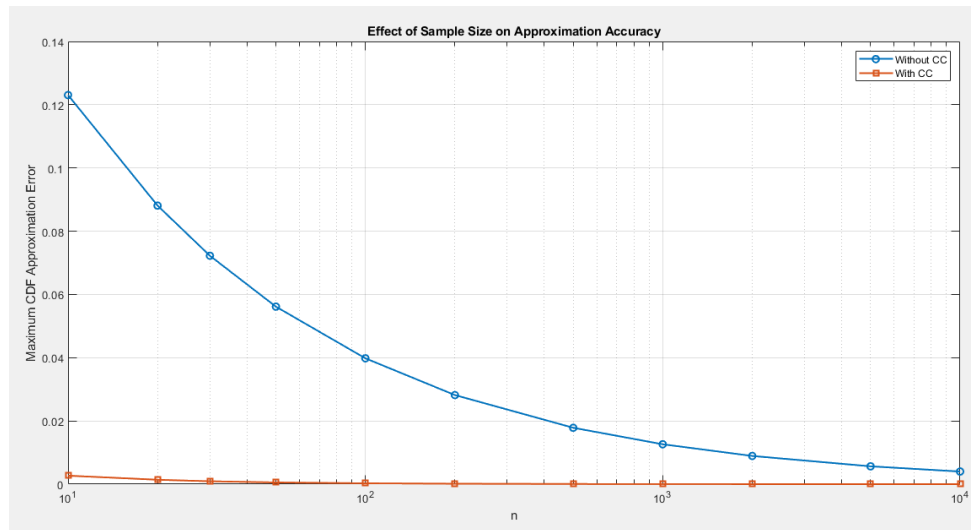


Figure 3. Shows the effect of sample size n on approximation accuracy at $P=0.5$

When $n=100$ was set, the results showed that the error is smaller near $P=0.5$ and increases as P approaches zero or one. $D_{Grid,CC}$ was 0.039795 at $P = 0.5$, while it was 0.23576 at $P = 0.01$ and $P = 0.99$. With the continuity correction, the values decreased to 0.00027259 at $P = 0.5$ and 0.058381 at $P = 0.01$, and 0.065834 at $P = 0.99$. This shows that the accuracy of the approximation is affected by the probability parameter as well as the sample size. The following figure illustrates the effect of the probability parameter when $n = 100$.

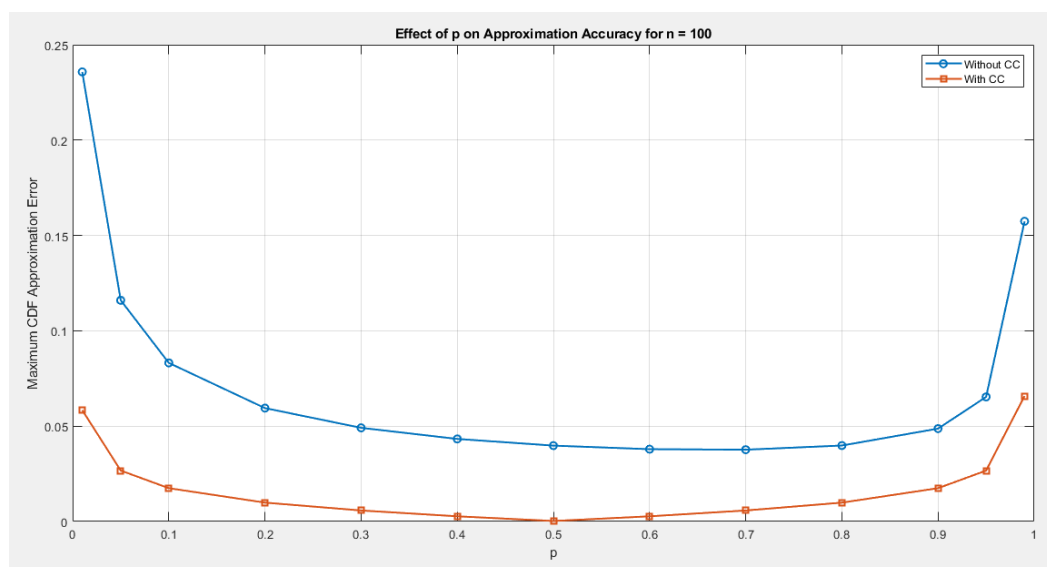


Figure 4. Effect of the probability parameter at $n=100$

Table 2. Shows the change in approximation accuracy with P for a fixed sample size of $n = 100$

P	nP	$n(1 - P)$	$ \gamma_1 $	D_{Grid} without continuity correction	D_{Grid} with continuity correction	Improvement (%)
0.01	1	99	0.98494	0.23576	0.058381	75.237
0.05	5	95	0.41295	0.116	0.026708	76.975
0.1	10	90	0.26667	0.083156	0.017474	78.986
0.2	20	80	0.15	0.059462	0.0098996	83.351
0.3	30	70	0.087287	0.049124	0.0057817	88.23
0.4	40	60	0.040825	0.043294	0.0027216	93.714
0.5	50	50	0	0.039795	0.00027259	99.315
0.6	60	40	0.040825	0.037925	0.0027216	92.824
0.7	70	30	0.087287	0.03766	0.0057817	84.648
0.8	80	20	0.15	0.039839	0.0098996	75.151
0.9	90	10	0.26667	0.04871	0.017474	64.126
0.95	95	5	0.41295	0.065339	0.026708	59.123
0.99	99	1	0.98494	0.15744	0.065834	58.185

The following figure shows the improvement achieved by applying the continuity correction when $n=100$.

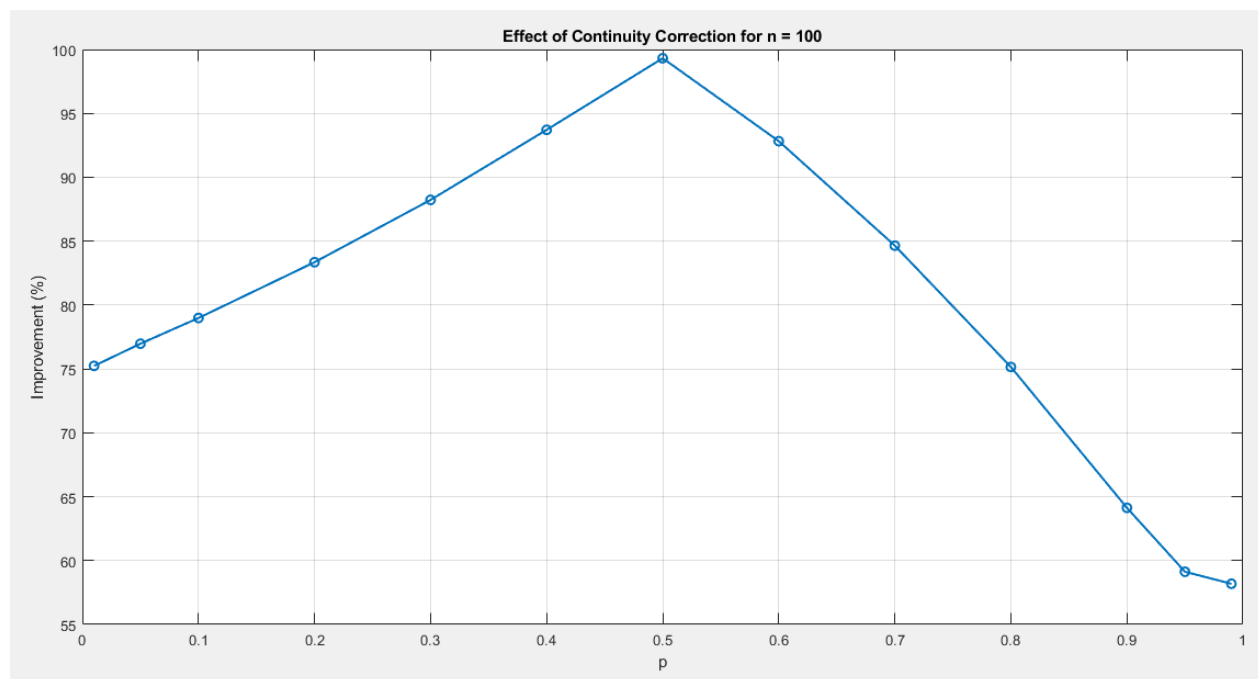


Figure 5. Percentage of improvement resulting from the continuity correction at $n=100$

The following heat maps provide a comprehensive picture of how the error varies as both the sample size n and the probability parameter P change.

The following figure is a heat map showing the maximum cumulative error without continuity correction.

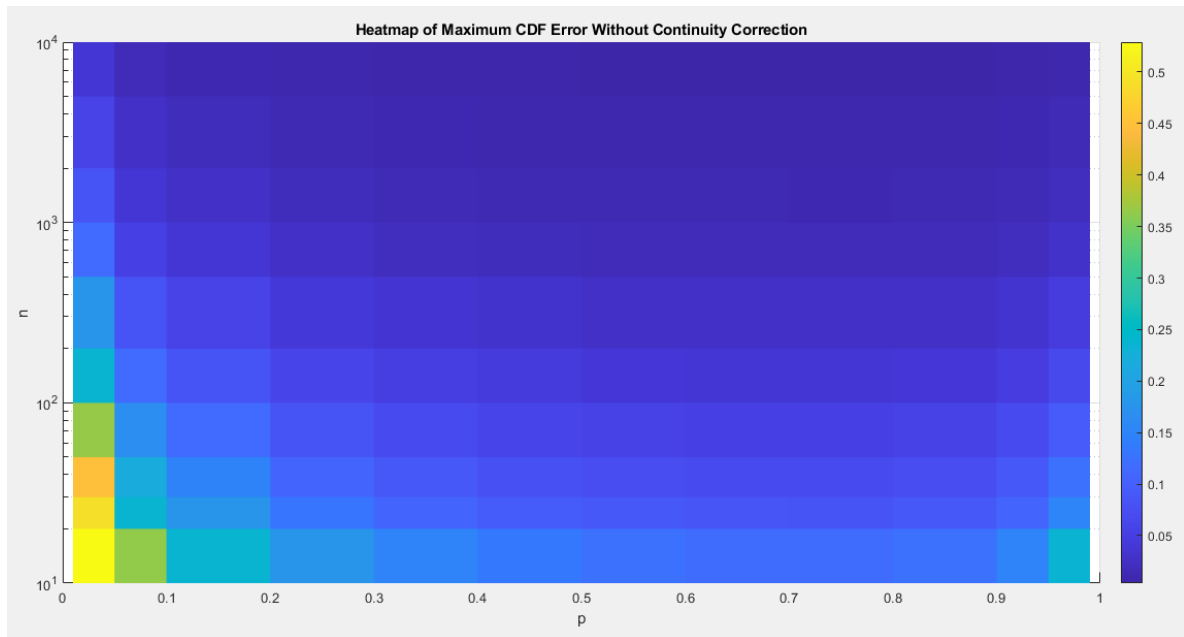


Figure 6. Heat map showing the maximum cumulative error without continuity correction

The following figure is a heat map showing the maximum cumulative error with continuity correction.

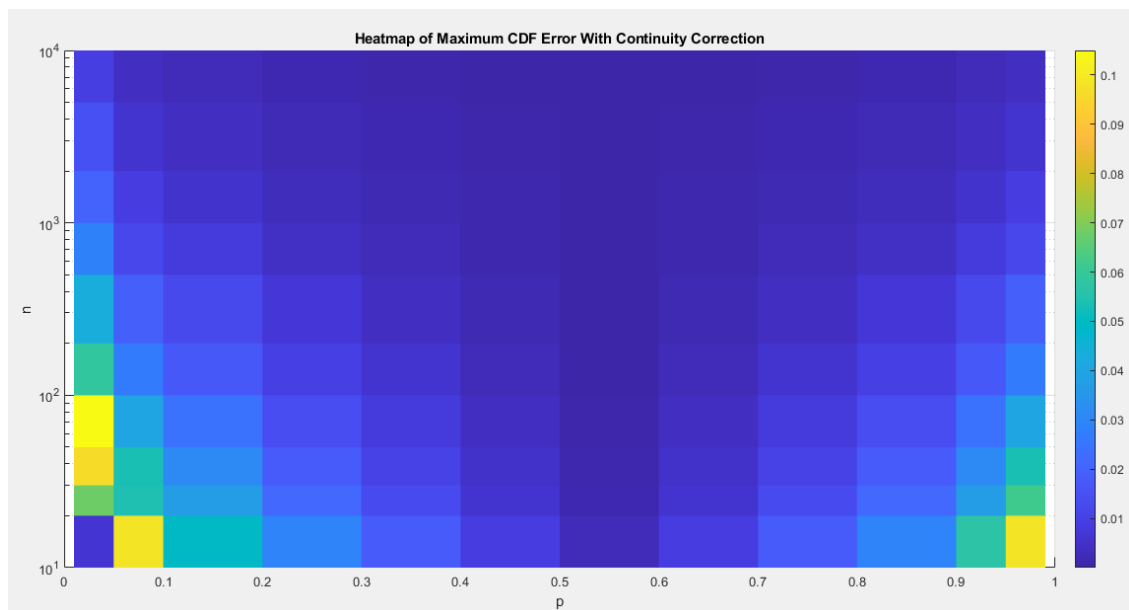


Figure 7. Heat map showing the maximum cumulative error with continuity correction

The correlation between the error and both nP and $n(1-P)$ was analyzed, and the results showed an inverse ordinal correlation between nP and the error, with Spearman's correlation coefficient for the approximation being (-0.928527) without correction and (-0.687835) with correction. The Spearman's correlation coefficient between $n(1-P)$ and the error was (-0.802084) without correction and (-0.692505) with correction. The following

table shows the Spearman's and Pearson's correlation coefficients between the study variables and the rounding error.

Table 3. Shows Spearman and Pearson coefficients between the study variables and the approximation error

variable	Pearson NC	<i>p</i> – value	Pearson CC	<i>p</i> – value	Spearman NC	<i>p</i> – value	Spearman CC	<i>p</i> – value
nP	0.319131	0.000102397	0.261742	0.00158992	-0.928527	1.52619×10^{-62}	-0.687835	2.32595×10^{-21}
$(1 - P)$	0.293917	0.000366756	0.261919	0.0015779	-0.802084	2.26287×10^{-33}	-0.692505	9.70181×10^{-22}
$ r_1 $	0.922252	4.60815×10^{-60}	0.735389	1.3452×10^{-25}	0.82912	1.9925×10^{-37}	0.98189	9.1434×10^{-104}

The following figure illustrates the relationship between cumulative error and nP

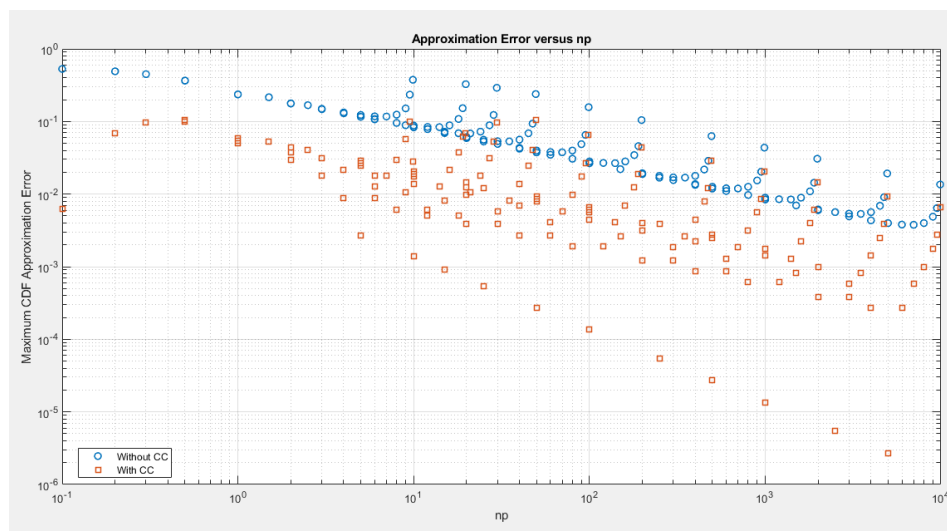


Figure 8. Shows the relationship between the cumulative error and nP

The following figure illustrates the relationship between the cumulative error and $n(1-P)$

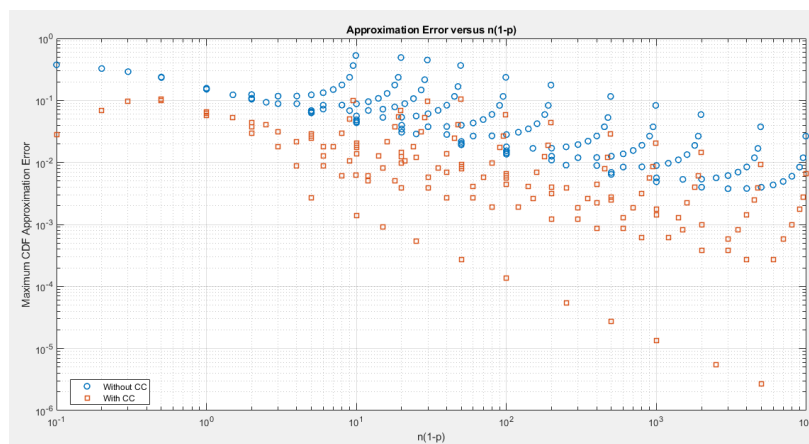


Figure 9. Shows the relationship between the cumulative error and $n(1-P)$

The absolute skewness of the binomial distribution was calculated based on $|\gamma_1|$, and the results showed a strong positive correlation between the skewness and the rounding error. The Pearson correlation coefficient was 0.922252 for the uncorrected approximation and 0.735389 for the corrected approximation, while the Spearman correlation coefficient was 0.82912 and 0.98189, respectively. These results indicate a strong correlation between the skewness and the rounding error within the scenarios studied. The following figure illustrates the relationship between the cumulative error and the absolute skewness $|\gamma_1|$.

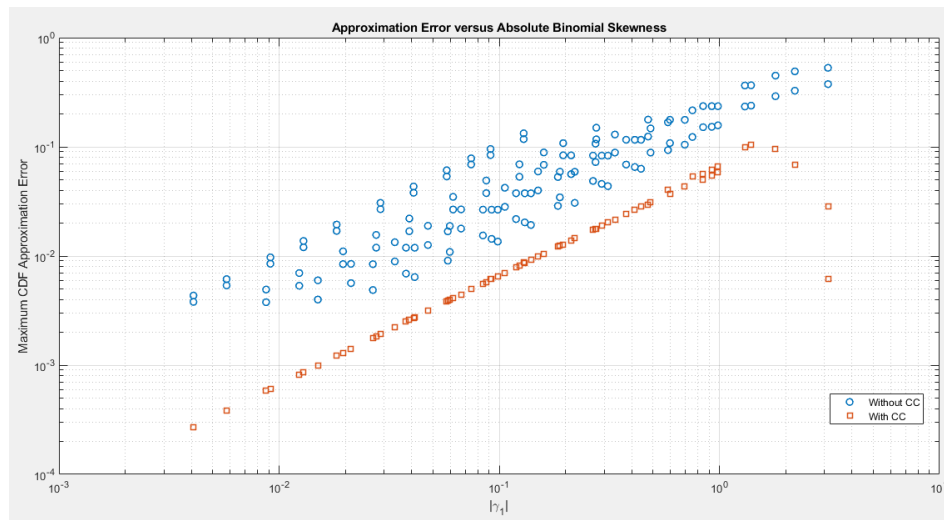


Figure 10. Shows the relationship between the cumulative error and the absolute skewness $|\gamma_1|$

The following figure illustrates the relationship between the absolute twist $|\gamma_1|$ and np

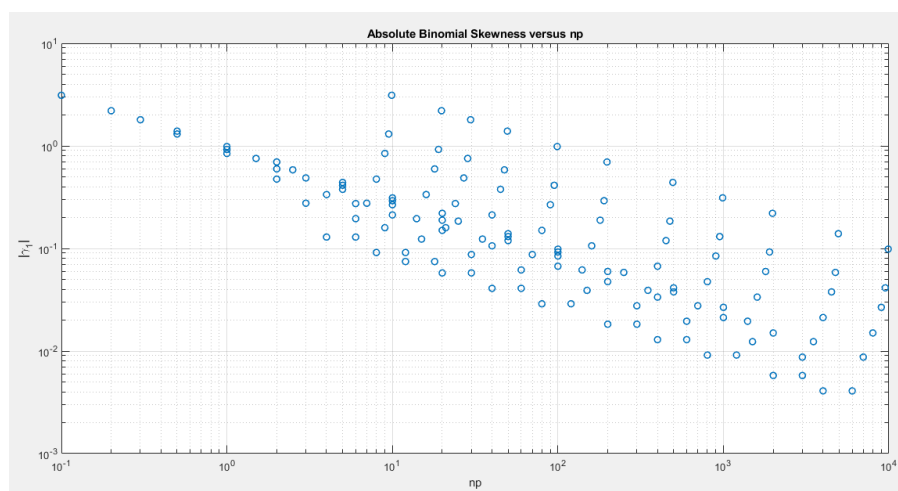


Figure 11. Shows the relationship between the absolute skewness $|\gamma_1|$ and np

The following figure shows the correlation coefficients between the error and each of nP , $n(1-P)$, and $|\gamma_1|$

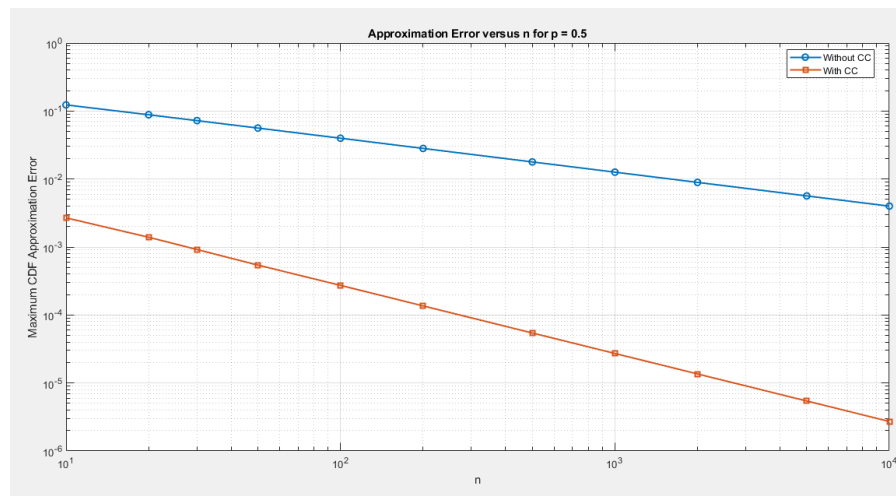


Figure 12. Shows the correlation coefficients between the error and each of nP , $n(1 - P)$, and $|\gamma_1|$

The following table shows the effect of sample size when $P = 0.5$

Table 4. Shows the effect of sample size at $P=0.5$

n	nP	$n(1 - P)$	D_{Grid} without continuity correction	D_{Grid} with continuity correction	Improvement (%)
10	5	5	0.12305	0.0026862	97.817
20	10	10	0.088099	0.0013909	98.421
30	15	15	0.072232	0.00091789	98.729
50	25	25	0.056138	0.0005417	99.035
100	50	50	0.039795	0.00027259	99.315
200	100	100	0.028174	0.00013603	99.517
500	250	250	0.017832	0.000054357	99.695
1000	500	500	0.012613	0.000027166	99.785
2000	1000	1000	0.0089195	0.000013576	99.848
5000	2500	2500	0.0056416	0.0000054306	99.904
10000	5000	5000	0.0039893	0.0000027152	99.932

Discussion

The results show that the accuracy of the normal approximation depends on both sample size and the shape of the binomial distribution. At $P = 0.5$, the maximum cumulative error decreased as n increased from 10 to 10,000, consistent with the theoretical basis of the approximation [9]. At a fixed sample size, however, errors were generally larger when P approached zero or one, where the binomial distribution

becomes more skewed. This pattern is supported by the positive correlations between absolute skewness and approximation error observed across the study scenarios. Continuity correction substantially improved accuracy by accounting for the difference between discrete binomial outcomes and a continuous normal distribution [8]: overall MAE decreased from 0.03625797 to 0.007330879, while RMSE decreased from 0.0690118 to 0.016216. The inverse correlations between error and both nP and $n(1 - P)$ further indicate why a large n alone is an insufficient guide to approximation quality. Monte Carlo estimates generally agreed with the exact binomial probabilities, providing an empirical check on the calculations; their random variation should be distinguished from the systematic error of the normal approximation. These findings support evaluating P , nP , $n(1 - P)$, skewness, and the location of the probability being estimated when deciding whether to use the normal approximation, particularly for probabilities near the ends of the distribution.

CONCLUSION

Fundamental Finding: The study showed that an increase in sample size n was associated with a decrease in the normal approximation error, while accuracy was greatest around $P = 0.5$ and declined as P approached zero or one. Continuity correction reduced MAE, RMSE, and maximum cumulative error, with an average improvement of 79.5393% for D_Grid. The error was inversely correlated with nP and $n(1-P)$ and positively correlated with absolute skewness. **Implication:** The accuracy of the normal approximation should be assessed using P , nP , $n(1-P)$, and the shape of the distribution alongside sample size, with continuity correction applied when estimating binomial probabilities. **Limitation:** The findings are based on the selected parameter values and evaluation points, while the Monte Carlo estimates retain random variation from the finite number of simulation iterations. **Future Research:** Further studies could examine additional parameter combinations and tail probabilities, compare alternative approximation methods, and use larger simulation runs to assess the stability of the findings.

REFERENCES

- [1] E. W. Weisstein, "Binomial distribution," *MathWorld*, 2002. [Online]. Available: <https://mathworld.wolfram.com>
- [2] E. Von Collani and K. Dräger, *Binomial Distribution Handbook for Scientists and Engineers*. New York, NY, USA: Springer Science & Business Media, 2001.
- [3] J. I. García-García, N. A. Fernández Coronado, E. H. Arredondo, and I. A. Imilpán Rivera, "The binomial distribution: Historical origin and evolution of its problem situations," *Mathematics*, vol. 10, no. 15, Art. no. 2680, 2022.
- [4] L. Jäntschi, "Formulas, algorithms and examples for binomial distributed data confidence interval calculation: Excess risk, relative risk and odds ratio," *Mathematics*, vol. 9, no. 19, Art. no. 2506, 2021.

- [5] A. Gianinetti, "Basic features of the analysis of germination data with generalized linear mixed models," *Data*, vol. 5, no. 1, Art. no. 6, 2020.
- [6] S. Wallis, "Binomial confidence intervals and contingency tests: Mathematical fundamentals and the evaluation of alternative methods," *Journal of Quantitative Linguistics*, vol. 20, no. 3, pp. 178–208, 2013.
- [7] M. J. Faddy and D. M. Smith, "Modeling the dependence between the number of trials and the success probability in binary trials," *Biometrics*, vol. 61, no. 4, pp. 1112–1114, 2005.
- [8] T. Emura and Y. T. Liao, "Critical review and comparison of continuity correction methods: The normal approximation to the binomial distribution," *Communications in Statistics – Simulation and Computation*, vol. 47, no. 8, pp. 2266–2285, 2018.
- [9] T. Rudas, "Normal approximations," in *Lectures on Categorical Data Analysis*, New York, NY, USA: Springer, pp. 41–56, 2018.
- [10] M. Thulin, "The cost of using exact confidence intervals for a binomial proportion," 2014.
- [11] H. J. Chang and M. C. Lee, "Applying computer simulation to analyze the normal approximation of binomial distribution," *Journal of Computers*, vol. 28, no. 5, pp. 116–131, 2017.
- [12] U. Michelucci, "Skewness, kurtosis, and modality," in *Statistics for Scientists: A Concise Guide for Data-Driven Research*, Cham, Switzerland: Springer Nature Switzerland, pp. 95–108, 2025.
- [13] G. Hatem, J. Zeidan, M. Goossens, and C. Moreira, "Normality testing methods and the importance of skewness and kurtosis in statistical analysis," *BAU Journal – Science and Technology*, vol. 3, no. 2, Art. no. 7, 2022.
- [14] U. Sarkar, S. Chakraborty, and K. S. Meel, "Assessing the quality of binomial samplers: A statistical distance framework," in *Proc. International Conference on Computer Aided Verification*, Cham, Switzerland: Springer Nature Switzerland, pp. 231–253, 2025.
- [15] A. Gasparini, "rsimsum: Summarise results from Monte Carlo simulation studies," *Journal of Open Source Software*, vol. 3, no. 26, Art. no. 739, 2018.

* **Alaa Farhan Ahmed Azawi (Corresponding Author)**
University of Tikrit, Iraq
Email: alaa.f@tu.edu.iq
