EMPIRICAL VERIFICATION OF THE CENTRAL LIMIT THEOREM BY COMPUTER SIMULATION

STEVEN LAROSA, HANNAH MANUEL, AND RAPHAEL MARIANO

1. Introduction

This project aims to verify the Central Limit Theorem for four different probability distributions by computer simulation. For a sequence of n *i.i.d.* random variables X_i, each with finite mean μ and finite variance σ^2, the theorem asserts that

$$\lim_{n\to\infty} P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right] = P\left[Z \leq z\right], \text{where } Z \sim Normal(0,1) \tag{1.1}$$

and where

$$\bar{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}.$$

is the sample mean. For a continuous distribution with range (a, b), the mean is defined as

$$\mu = \frac{1}{b-a}\int_a^b f(x)\,dx,$$

and for a discrete distribution,

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i$$

For a continuous distribution the variance is defined as

$$\sigma^2 = \int (x-\mu)^2 f(x)dx,$$

and for a discrete distribution,

$$\sigma^2 = \sum_{i=1}^{n}(x_i-\mu)^2 f_i$$

In each case, $f(x)$ represents the probability density function (pdf) of the distributions. The pdf gives the probability of each outcome in the sample space of the distribution. The four distributions used are the Poisson distribution, the binomial distribution, the exponential distribution, and the Irwin-Hall distribution. For each distribution, 1000 samples each of size $n = 40$ were drawn using algorithms that employ the random number function in Mathematica. These sample averages were standardized using the equation

$$X^* = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{1.2}$$

The resulting distributions were then compared to the standard normal distribution by plotting the histograms and comparing visually, by comparison of results to the numerical value of certain z-values on the standard normal distribution, and by the Chi-square goodness-of-fit test.

Definition 1.1. A normal distribution with mean μ and variance σ^2, written as $Normal(\mu, \sigma^2)$, is defined as

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty \tag{1.3}$$

A variable $X \sim Normal(\mu, \sigma^2)$ is called a normal random variable. We also say that X is normally distributed. [2, p.37]

Definition 1.2. The probability mass function of a discrete random variable X, denoted as $p(a)$, returns the probability that $X = a$. [2, p.27]

Definition 1.3. The probability density function $f(x)$ of a continuous random variable X gives the probability $P[a \leq X \leq b]$ by evaluating $\int_a^b f(x)\, dx$. [2, p.34]

Definition 1.4. The moment generating function $\phi(t)$ of a random variable X is defined for all values of t by the expected value for e^{tX}.

$$\begin{aligned} \phi(t) &= E\left[e^{tx}\right] \\ &= \begin{cases} \sum_x e^{tX} p(x) & \text{, if } X \text{ is discrete} \\ \int_{-\infty}^{\infty} e^{tx}\, f(x)\, dx & \text{, if } X \text{ is continuous.} \end{cases} \end{aligned} \tag{1.4}$$

[2, p.64]

Two notable properties of moment generating functions are:

(1) The moment generating function of the sum of independent random variables is just the product of the individual moment generating functions.[2, p.68]
(2) The moment generating function uniquely determines the distribution.[2, p.69]

The cumulative distribution function (denoted cdf) of the random variable X is defined by $F(b) = P[X \leq b]$ for any real number b in $(-\infty, \infty)$.

A variable $X \sim Normal(0,1)$ is a standard normal random variable. Its cumulative distribution function is $P[Z \leq z] = F(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx$. Its moment generating function is $\phi(t) = e^{t^2/2}$.

1.1. **Central Limit Theorem.**

Theorem 1.5 (The Central Limit Theorem)**.** *Let $X_1, X_2, \ldots$ be a sequence of independent, identically distributed random variables, each with mean μ and variance σ^2. Then the distribution of*

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

approaches the standard normal curve as $n \to \infty$. That is, as $n \to \infty$:

$$P\left\{\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq z\right\} \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2}\, dx \tag{1.5}$$

[2, 79]

The following heuristic proof of the central limit theorem is from Ross [2, p.82-83].

Proof. Suppose that every X_i in $X_1, X_2, \ldots, X_n$ has mean 0 and variance 1, and let $E\left[e^{tX}\right]$ denote their common moment generating function. Then, the expression $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ will have a moment generating function of

$$\begin{aligned}
\phi(t) &= E\left[\exp\left\{\mathrm{t}\left(\frac{X_1 + \cdots + X_n}{\sqrt{n}}\right)\right\}\right], \text{ by } (1.4) \\
&= E\left[e^{tX_1/\sqrt{n}} e^{tX_2/\sqrt{n}} \cdots e^{tX_n/\sqrt{n}}\right]
\end{aligned}$$

Since each X_i in $X_1, X_2, \ldots, X_n$ is independent and has a common moment generating function, we have that

$$E\left[e^{tX_1/\sqrt{n}} e^{tX_2/\sqrt{n}} \cdots e^{tX_n/\sqrt{n}}\right] = \left(E\left[e^{tX/\sqrt{n}}\right]\right)^n \tag{1.6}$$

We then obtain from the Taylor series expansion of $e^{tX/\sqrt{n}}$ for large values of n that

$$e^{tX/\sqrt{n}} \approx 1 + \frac{tX}{\sqrt{n}} + \frac{t^2X^2}{2n}$$

Since each X_i has mean 0 and variance 1, taking expectations for when n is large we get

$$\begin{aligned} E\left[e^{tX/\sqrt{n}}\right] &\approx 1+\frac{tE[X]}{\sqrt{n}}+\frac{t^2E[X^2]}{2n} \\ &= 1+\frac{t^2}{2n} \end{aligned} \tag{1.7}$$

By combining (1.6) and (1.7), we get

$$E\left[\exp\left\{\text{t}\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] \approx \left(1+\frac{t^2}{2n}\right)^n \tag{1.8}$$

By taking the limit of (1.8) as n approaches ∞, we get:

$$\lim_{n\to\infty} E\left[\exp\left\{\text{t}\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] = \lim_{n\to\infty}\left(1+\frac{t^2}{2n}\right)^n \tag{1.9}$$

which is in the indeterminate form of 1^∞.

To solve this expression, we let

$$\begin{aligned} y &= \left(1+\frac{t^2}{2n}\right)^n, \text{ then} \\ ln(y) &= ln\left[\left(1+\frac{t^2}{2n}\right)^n\right] \\ &= n*ln\left(1+\frac{t^2}{2n}\right) \\ \lim_{n\to\infty} ln(y) &= \lim_{n\to\infty} n*ln\left(1+\frac{t^2}{2n}\right) \end{aligned} \tag{1.10}$$

which is in the indeterminate form $\infty \cdot 0$. Rearranging the terms allow us to apply L'Hospital's Rule:

$$\begin{aligned}
\lim_{n\to\infty} n * ln\left(1+\frac{t^2}{2n}\right) &= \lim_{n\to\infty} \frac{ln\left(1+\frac{t^2}{2n}\right)}{\frac{1}{n}} \quad \left(\text{in the indeterminate form } \frac{0}{0}\right) \\
&= \lim_{n\to\infty} \frac{\left(\frac{1}{1+\frac{t^2}{2n}}\right)\left(\frac{t^2}{2}\right)\left(\frac{-1}{n^2}\right)}{\frac{-1}{n^2}} \quad \text{(by L'Hospital's Rule)} \\
&= \lim_{n\to\infty} \left(\frac{1}{1+\frac{t^2}{2n}}\right)\left(\frac{t^2}{2}\right) \\
&= \left(\frac{1}{1+0}\right)\left(\frac{t^2}{2}\right) \\
&= \frac{t^2}{2} \qquad (1.11)
\end{aligned}$$

Combining equations (1.9), (1.10) and (1.11), we get:

$$y = e^{ln(y)} = \lim_{n\to\infty} E\left[\exp\left\{\mathrm{t}\left(\frac{X_1+\cdots+X_n}{\sqrt{n}}\right)\right\}\right] = e^{t^2/2}, \qquad (1.12)$$

which is the moment generating function of a standard normal random variable. Thus, the moment generating function of $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the moment generating function of a standard normal random variable with mean 0 and variance 1. Using the uniqueness of the moment generating functions, we can say that the distribution function of the random variable $\frac{X_1+\cdots+X_n}{\sqrt{n}}$ converges to the standard normal distribution function. [2, p.82-83]

In the case of random variable X_i having mean μ and variance σ^2, the random variable $U_i = \frac{X_i-\mu}{\sigma}$ has a mean of 0 and variance 1. The above proof can then be applied. □

Remark 1.6. Another version of the central limit theorem states that:

Theorem 1.7 (The Central Limit Theorem*)**.** *Let X be a random variable that follows a distribution with finite mean μ and finite variance σ^2, and let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from this distribution. Then the sample average*

$$\bar{X} \equiv \frac{X_1+X_2+\ldots+X_n}{n}, \qquad (1.13)$$

when the sample size n is large, follows a distribution which is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$. More precisely, it asserts the following asymptotic result:

$$\lim_{n\to\infty} P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right] \le z] = P[Z \le z]$$

Proof.

$$\begin{aligned}
\lim_{n\to\infty} P\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right] &= \lim_{n\to\infty} P\left[\frac{\frac{X_1+X_2+\ldots+X_n}{n}-\mu}{\sigma/\sqrt{n}}\right], \text{ by (1.13)} \\
&= \lim_{n\to\infty} P\left[\frac{\frac{X_1+X_2+\ldots+X_n-n\mu}{n}}{\sigma/\sqrt{n}}\right] \\
&= \lim_{n\to\infty} P\left[\frac{X_1+X_2+\ldots+X_n-n\mu}{(\sigma/\sqrt{n})\,(n)}\right] \\
&= \lim_{n\to\infty} P\left[\frac{X_1+X_2+\ldots+X_n-n\mu}{\sigma\sqrt{n}}\right] \\
&= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{z} e^{-x^2/2}\,dx, \text{ by (1.5)} \\
&= P[Z \le z] \quad \text{cdf of Normal(0,1)}
\end{aligned}$$

We have that:

$$\begin{aligned}
E[\bar{X}] &= \frac{1}{n}\sum_{i=1}^{m} E[X_i] \\
&= \mu \qquad [2, \text{p.55}]
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= \left(\frac{1}{n}\right)^2 \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \left(\frac{1}{n}\right)^2 \sum_{i=1}^{n} \mathrm{Var}(X_i) \\
&= \frac{\sigma^2}{n} \qquad [2, \text{p.55}]
\end{aligned}$$

$\square$

2. Method

For each distribution, the 1000 sample means were tabulated and standardized using (1.2). First, these means were graphed on a histogram to compare their distribution pictorially with a standard normal distribution. Psuedo-probabilites were then calculated for several values of z in the following manner:

$$P_{psuedo}\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right] = \frac{number\ of\ \bar{X} \leq z}{1000} \tag{2.1}$$

These values were then compared to the actual values given by the CDF of the standard normal distribution using z as the upper limit of integration. Finally, the standardized means were verified using the Chi-square goodness-of-fit test.

2.1. **Inverse Transformation Method.**

Proposition 2.1. *To simulate a random variable with a continuous distribution function F using the inverse transformation method, set the random variable*

$$X = F^{-1}[U] \tag{2.2}$$

where U is a $U(0,1)$ random variable. The random variable X has the same distribution function F.

Proof.

$$\begin{aligned} F_X(a) &= P[X \leq a] \\ &= P[F^{-1}(U) \leq a] \end{aligned}$$

Since F is a monotonically increasing function,

$$\begin{aligned} F_X(a) &= P[U \leq F(a)] \\ &= F(a) \end{aligned}$$

□

2.2. **The Chi-Square Goodness-of-Fit Test.** The chi-square goodness-of-fit test statistically determines whether a set of data follows a hypothesized distribution. Intuitively, it determines how distant observed values are from expected values.

There are two assumptions for the goodness-of-fit test:

(1) The data are obtained from a random sample; and
(2) The expected frequency for each category is 5 or more.

If the data satisfies the assumptions, a hypothesis can now be formed.

Definition 2.2. A null hypothesis, denoted as H_0, is a statistical hypothesis that states that there is no difference between two distributions.

Definition 2.3. An alternative hypothesis, denoted as H_1, is a statistical hypothesis that states the existence of a difference between distributions[1] .

The null hypothesis being tested is that there is no difference between the distribution of sample means and a standard normal distribution. Now that a hypothesis is present, the data or the distribution of data is separated into categories of sufficient size. Too many categories can result with expected frequencies being less than 5, dissatisfying assumption (2). Two parameters are then identified: degrees of freedom, and α. The parameter degrees of freedom is one less than the number of categories. The parameter α is a value used in statistics that corresponds to the probability for error. This parameter, as a type of probability, is a number between 0 and 1. We choose $\alpha = 0.05$.

A test statistic is then determined using the formula:

$$\chi^{2^*} = \sum \frac{(O-E)^2}{E}, \text{ where} \tag{2.3}$$

$$\begin{aligned} O = &\ \text{ observed frequency;} \quad \text{and} \\ E = &\ \text{ expected frequency.} \end{aligned}$$

The value of χ^{2^*} derived from equation (2.3) is then compared to a critical value. This value is determined by the chi-squared distribution at the chosen α and the degrees of freedom. If χ^{2^*} is less than the critical value, then there is not enough evidence to reject the hypothesis that a distribution follows the predetermined pattern.[2] [1, p.585-592]

When the pattern being compared is the normal distribution, the expected values used should be

$$E = n \int_{z_{min}}^{z_{max}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \text{ where} \tag{2.4}$$

---

[1]These definitions use the word *distributions, but these definitions can extend to include terms such aas parameters and randomness.*

[2]In statistics, one cannot conclude that a hypothesis is true. One can only reject or fail to reject a hypothesis based on the data from the goodness-of-fit test.

$$\begin{aligned} n &= \text{total number of sample averages,} \\ z_{max} &= \text{z-value of upper limit of category; and} \\ z_{min} &= \text{z-value of lower limit of category.} \end{aligned}$$

3. The Four Distributions

3.1. **Poisson Distribution.** The Poisson distribution gives the probability that a number of events k will occur in a set period of time with the parameter λ being the expected number of events to occur in the given period of time. A discrete random variable X follows the $Poisson(\lambda)$ distribution, $\lambda > 0$, if its probability mass function is given by

$$f(k) = \frac{\lambda^k e^{-\lambda}}{k!}, \; where \; k \; \in \{0, 1, 2, ...\}.$$

The expected value and variance of a random variable $X \sim Poisson(\lambda)$ can be determined by

$$E(X) = \lambda \tag{3.1}$$

$$Var(X) = \lambda \tag{3.2}$$

A Poisson random variable N can be simulated by:

$$N = min\{n : \prod_{i=1}^{n} U_i < e^{-\lambda}\} - 1 \;, \; where \; U_i \sim U(0, 1).$$

The following histogram illustrates a $Poisson(4)$ distribution simulated with the above method.

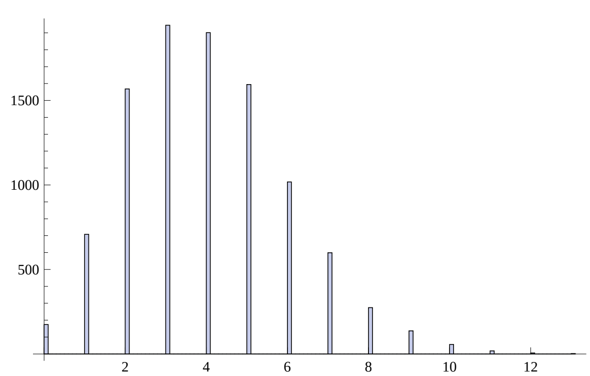


Figure 1. Histogram showing the distribution of simulated random variable $X \sim Poisson(4)$.

3.2. **Binomial Distribution.** A binomial distribution, denoted as $Bin(n,p)$, depicts n independent discrete random variables, each with probability p of success and probability $(1-p)$ of failure. The probability mass function of a binomial random variable having parameters (n,p) is given by

$$P[X=x] = \binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}$$

The expected value and variance of a binomial distribution can be determined by

$$E(X) = np \tag{3.3}$$

$$Var(X) = np(1-p) \tag{3.4}$$

To simulate the binomial distribution, we must introduce the **Bernoulli distribution**. The Bernoulli distribution depicts a single event with probability p of success and $(1-p)$ of failure. It can be described as the case of the binomial distribution with parameters $(1,p)$. A random variable $X \sim Bin(n,p)$ can however be defined as the number of successes from n repetitions of $Bernoulli(p)$. The simulation algorithm is based on this principle. Ross [2, p.686-687] proposes the following steps to simulate a binomial distribution with events X_k:

(1) Let $\alpha = \frac{1}{p}$; $\beta = \frac{1}{1-p}$
(2) Set a counter k to 0.
(3) Generate a uniform random number from $U(0,1)$.
(4) If $k=n$, stop. Otherwise, reset k to equal $k+1$
(5) If $U \leq p$, then $X_k = 1$ and reset U to αU. If $U > k$, then $X_k = 0$ and reset U to $\beta(U-p)$. Return to Step 4.

This algorithm uses $U(0,1)$ only once. The randomness of the consequent values is based on the uniformity of $(0,p)$ for the event of a success and the uniformity of $(p,1)$, due to the use of α and β. To find out how many successes have occured, we only need to count the number of instances where $X_k = 1$.

The following histogram illustrates a $Bin(100, 0.3)$ distribution simulated with the above method.

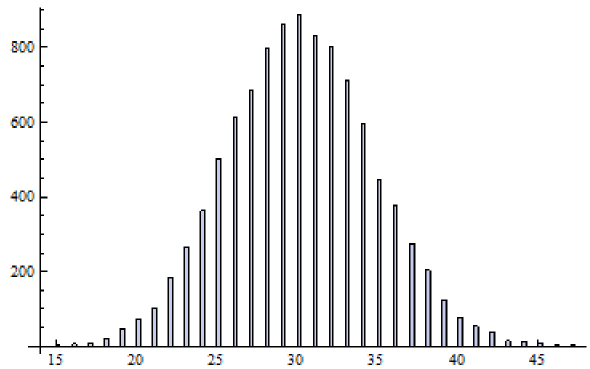


FIGURE 2. Histogram showing the distribution of simulated random variable $X \sim Bin(100, 0.3)$.

3.3. **Irwin-Hall Distribution.** The Irwin-Hall distribution is a continuous probability distribution of the sum of k independent uniformly distributed random variables on the interval $(0, 1)$. Though the distribution functions themselves become quite complicated for large values of n, simulation is fairly simple due to the definition of the distribution. We consider the case of the sum of $k = 4$ random variables. To simulate, using Mathematica, a table of 1000 means of samples of size $n = 40$ were generated. Each element in each sample was the sum of four different random variables generated by Mathematica from $U(0, 1)$. The following is the probability density function for the case of $k = 4$.

$$f_X(x) = \begin{cases} \frac{1}{6}x^3 & \text{, when } 0 \leq x \leq 1 \\ \frac{1}{6}(-3x^3 + 12x^2 - 12x + 4) & \text{, when } 1 \leq x \leq 2 \\ \frac{1}{6}(3x^3 - 24x^2 + 60x - 44) & \text{, when } 2 \leq x \leq 3 \\ \frac{1}{6}(-x^3 + 12x^2 - 48x + 64) & \text{, when } 3 \leq x \leq 4 \end{cases}$$

The expected value and variance of the Irwin-Hall distribution of the sum of k random variables from $U(0, 1)$ can be determined by

$$E(X) = k/2 \tag{3.5}$$
$$Var(X) = k/12 \tag{3.6}$$

The following histogram illustrates an $Irwin(4)$ distribution simulated with the above method.

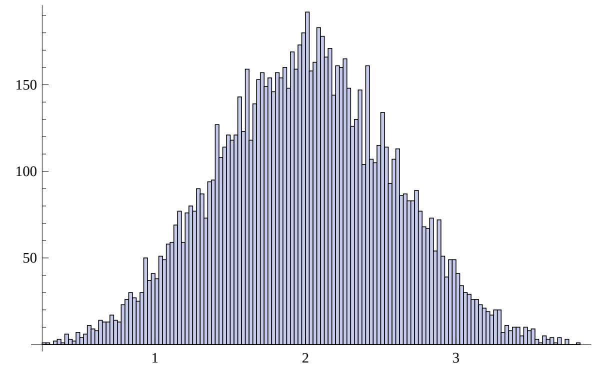


FIGURE 3. Histogram showing the distribution of simulated random variable $X \sim Irwin(4)$.

3.4. **Exponential Distribution.** The exponential distribution expresses the time between continuously and independently occurring events of a random process with an average rate of λ. A continuous random variable X follows the exponential distribution with parameter λ, $\lambda > 0$, if its cumulative distribution function is given by

$$F(x) = \begin{cases} 1 - e^{-\lambda x} & \text{, when } x \geq 0, \\ 0 & \text{, when } x < 0. \end{cases}$$

The expected value and variance of a random variable $X \sim Exp(\lambda)$ can be determined by

$$E(X) = \frac{1}{\lambda} \tag{3.7}$$

$$Var(X) = \frac{1}{\lambda^2} \tag{3.8}$$

The following simulates a random variable $X \sim Exp(\lambda)$ using the inverse transformation method (2.2) by generating a $U(0,1)$ random variable:

$$X = F^{-1}(U) = \frac{\ln(1-U)}{-\lambda}.$$

The following histogram illustrates an $Exp(4)$ distribution simulated with the above method.

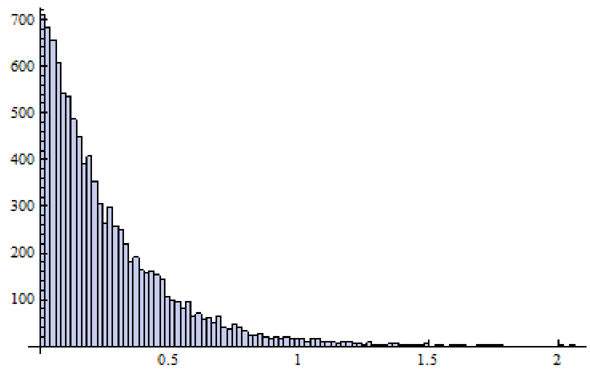
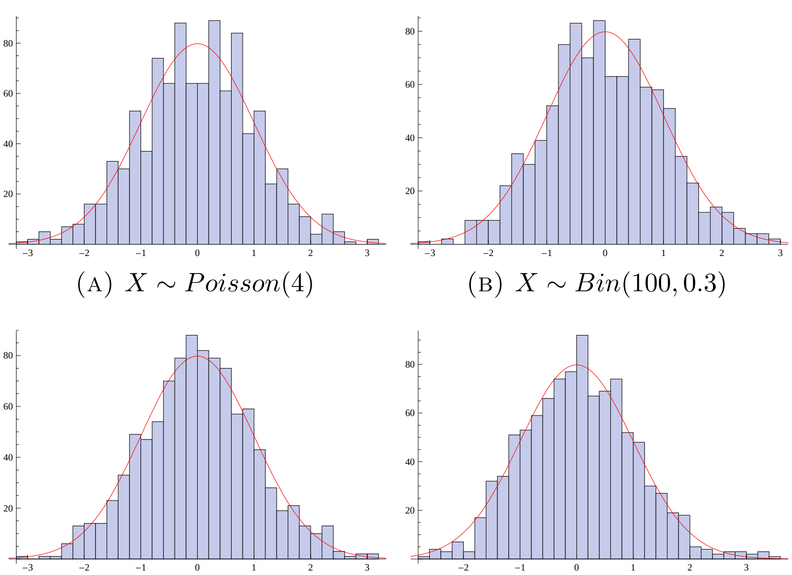


FIGURE 4. Histogram showing the distribution of simulated random variable $X \sim Exp(4)$.

4. Data and Results

4.1. **Comparing Histograms.** The following histograms show the distribution of the 1000 standardized simulated sample averages for each of the four distributions. The standard normal curve is plotted on each histogram to pictorially compare each distribution of sample averages with the standard normal distribution.



(A) $X \sim Poisson(4)$

(B) $X \sim Bin(100, 0.3)$

(C) $X \sim Irwin(4)$

(D) $X \sim Exp(4)$

FIGURE 5. Histograms showing the standardized distribution of 1000 sample averages of a simulated random variable X of sample size $n = 40$.

The distribution of the sample averages for each of the simulated distributions closely fits the standard normal distribution, supporting the Central Limit Theorem.

4.2. **Comparing Probabilities.** The following table compares the pseudo-probabilities, calculated with (2.1), of each of the four simulated standardized distributions of sample averages with the actual $P[Z \leq z]$, where $Z \sim Normal(0,1)$, for values of z, where $-2 \leq z \leq 2$ and has a step size of 0.4. The $P[Z \leq z]$ is given by the cumulative distribution function of the standard normal distribution using z as the upper limit of integration. The percent error for a certain value of z is given by

$$\%error = \frac{|P_{psuedo}\left[\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \leq z\right] - P[Z \leq z]|}{P[Z \leq z]} \times 100. \tag{4.1}$$

The percent error between each pseudo-probability and the corresponding value of $P[Z \leq z]$ for the standard normal distribution for a certain value of z is given in parentheses following the pseudo-probability for that z value.

	Pseudo-probabilities from Simulation				Actual
z	$Poisson(4)$	$Bin(100, 0.3)$	$Irwin(4)$	$Exp(4)$	$Normal(0,1)$
-2.0	0.025 (9.89%)	0.021 (7.69%)	0.022 (3.30%)	0.015 (34.07%)	0.023
-1.6	0.057 (4.02%)	0.052 (5.11%)	0.050 (8.76%)	0.035 (36.13%)	0.055
-1.2	0.120 (4.28%)	0.116 (0.81%)	0.106 (7.88%)	0.101 (12.23%)	0.115
-0.8	0.210 (0.88%)	0.207 (2.29%)	0.202 (4.65%)	0.205 (3.24%)	0.212
-0.4	0.348 (0.99%)	0.365 (5.93%)	0.326 (5.39%)	0.330 (4.23%)	0.345
0.0	0.500 (2.22%)	0.519 (3.80%)	0.493 (1.40%)	0.481 (3.80%)	0.500
0.4	0.653 (0.37%)	0.645 (1.59%)	0.654 (0.22%)	0.640 (2.35%)	0.655
0.8	0.798 (1.25%)	0.781 (0.91%)	0.786 (0.27%)	0.783 (0.65%)	0.788
1.2	0.895 (1.14%)	0.890 (0.57%)	0.888 (0.35%)	0.883 (0.22%)	0.885
1.6	0.949 (0.40%)	0.946 (0.08%)	0.935 (1.08%)	0.940 (0.55%)	0.945
2.0	0.976 (0.13%)	0.972 (0.54%)	0.969 (0.84%)	0.977 (0.03%)	0.977

FIGURE 6. Table of pseudo-probabilities for the distribution of the standardized sample averages for the four simulated distributions and the $P[Z \leq z]$ for the standard normal distribution evaluated at certain z-values. (Percent error)

The psuedo-probabilites of the simulated distributions are close to the actual values $P[Z \leq z]$ evaluated at the given z values, which supports the Central Limit Theorem. In 41 out of the 44 pseudo-probabilities, the percent error was less than 10 percent.

4.3. **Chi-Square Test Statistics.** The null hypothesis being tested is that the distribution of sample means illustrates a standard normal distribution.

To simplify the method, each distribution was standardized before being separated into 6 intervals:

$$(-\infty, -2], (-2, -1], (-1, 0], (0, 1], (1, 2], (2, \infty)$$

The number of sample means that lie in each interval was counted and used as the observed frequencies for (2.3). Formula (2.4) was used to determine the expected frequencies.

Since 6 intervals were used, our critical value according to the table of values in Bluman[1, p.772] is 11.071 at $\alpha = 0.05$ and degrees of freedom = 5. The test statistics for each distribution are shown in the following table:

Distribution	$Poisson(4)$	$Bin(100, 0.3)$	$Irwin(4)$	$Exp(4)$
Test Statistic	1.998	4.273	4.487	3.841

Figure 7. Table of test statistics acquired for the Chi-square goodness-of-fit test

Since all of the test statistics are less than the critical value of 11.071, we fail to reject our null hypothesis. There is not enough data to exhibit a difference between the distribution of sample means for all 4 distributions and the standard normal distribution. This statistically verifies the central limit theorem.

5. Conclusions

For each distribution, the results of the simulation and analysis supported the claim of the Central Limit Theorem. However, the theorem states the asymptotic result of a limit as the sample size n goes to infinity. While we do not show any results generated by varying n, trials were taken with n values of 400 and 4000. These simulations with the greater n values did not show any significant difference in accuracy of the X^* value when applied to (1.7). In some cases, the higher n value yielded a less accurate simulation. This is most likely due to the nature of the simulation, and the fact that 40 is already a large enough sample size for a good simulation.

Acknowledgements

The authors would like to thank Dr. Walfredo Javier for posing the problem and for his instruction; Tyler Moss, their graduate student mentor, for his time, guidance, and direction; and Dr. Mark Davidson for overseeing the SMILE summer program at Louisiana State University, in which the authors participated. The authors were funded by the National Science Foundation (NSF) through the Vertical Integration of Research and Education (VIGRE) program, which is part of the NSF Enhancing the Mathematical Sciences Workforce in the 21st Century.

References

[1] Bluman, A.G., *Elementary Statistics: a Step by Step Approach*, 6th ed., McGraw-Hill Companies, New York, 2007.

[2] Ross, S.M., *Introduction to Production Models*, 9th ed., Academic Press, Berkeley, 2007.