Analysis of the Distribution of *Ones* and *Zeros* in the Binary Expansion of π

Will Schenk

Abstract

This report analyzes the distribution and frequency of ones and zeros in the binary expansion of π. I hypothesize that each digit in the binary sequence is equally likely to be a one or zero. Using computational techniques, the first $100,000$ decimal digits of $\pi - 3$ are converted to binary and divided into non-overlapping segments. Their frequency distributions are compared against the theoretical **binomial distribution**. Additionally, the **Law of Large Numbers** and the **Central Limit Theorem** are applied to shed light on the nature of π's binary expansion.

1 Conversion of π to Binary

When represented in binary, π is expected to be non-terminating and non-repeating.

$$\text{Decimal of } \pi = \pi - 3$$

Due to computational constraints in MATLAB, John Machin's formula is used to convert the computation of π into a string array. This string is then converted to a binary sequence.

2 Frequency Distribution Analysis

Divide the binary sequence into non-overlapping segments of 10 digits and tally the number of ones in each segment. From this, we derive an empirical frequency distribution, which we then compare to a theoretical binomial distribution with parameters $n = 10$ *(number of trials)* and $p = 0.5$ *(probability of a one)*. The binomial probability mass function is:

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

Where $P(X = k)$ is the probability of observing k ones in a 10-digit sequence.

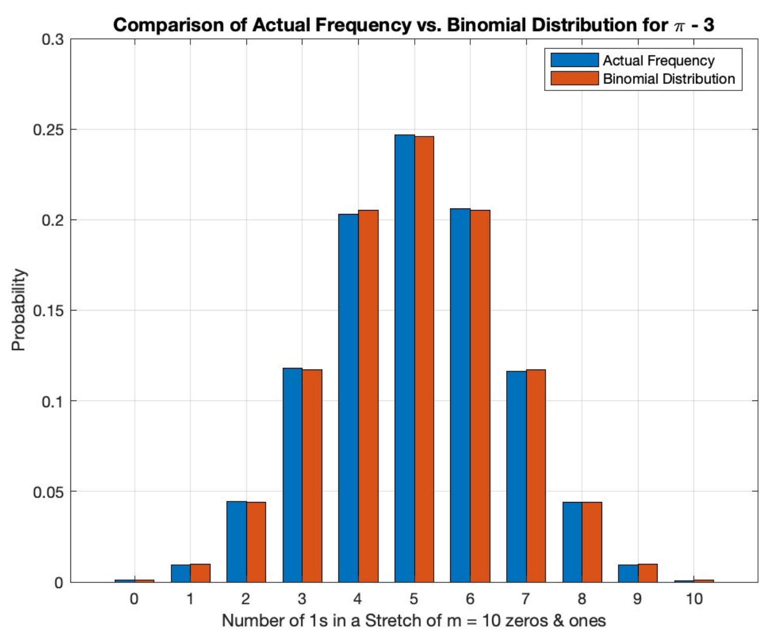

Figure 1: Actual Versus Binomial Distribution Bar Plot

3 Law of Large Numbers

The Law of Large Numbers states that as the sample size grows, its mean converges to the population mean. For a truly random binary sequence representing the digits of π, the sample mean of the sequence, denoted as $\bar{X}_n$, can be formulated as:

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

where each X_i is a random variable representing the i^{th} binary digit of π. Here, $X_i = 1$ with probability $p = 0.5$ and $X_i = 0$ with probability $1 - p = 0.5$. As n grows large, according to the Law of Large Numbers, $\bar{X}_n$ converges to the expected value $E[X]$:

$$E[X] = \sum_i x_i P(X = x_i) = 1(0.5) + 0(0.5) = 0.5$$

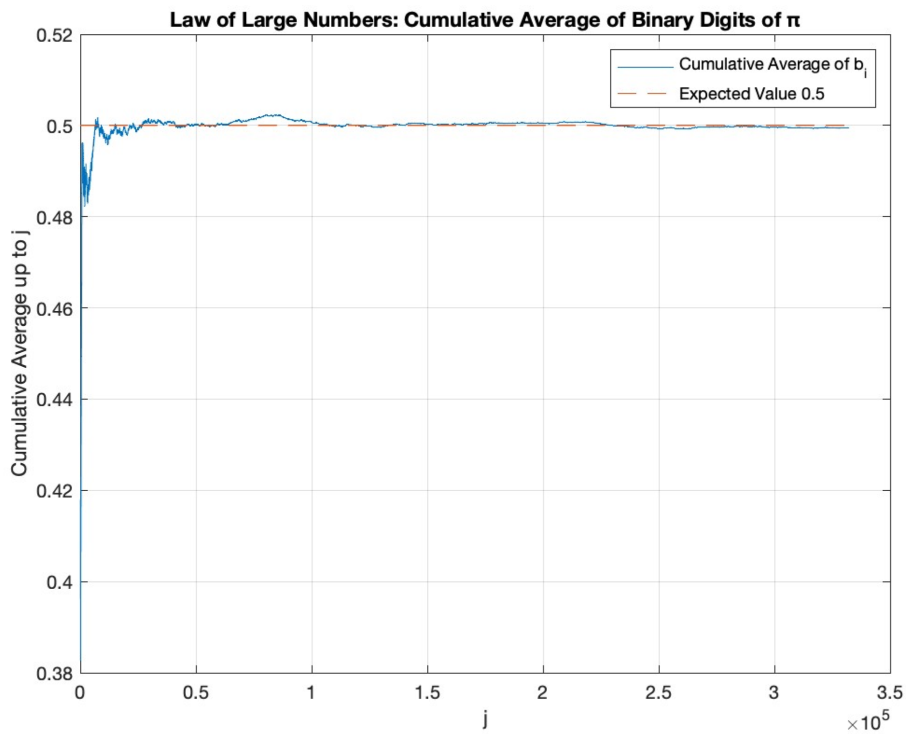

Figure 2: Cumulative average of ones in the binary sequence of π as the sample size increases.

As we move to the right, signifying larger sample sizes, the curve begins to stabilize closely around the value of 0.5. This stabilization is a testament to the Law of Large Numbers, indicating that with increasing sample size, the average value approaches the true population mean, which in this case is 0.5.

4 Central Limit Theorem

The Central Limit Theorem (CLT) states that, given a set of independent, identically distributed random variables, the distribution of their sum (or average) will approximate a normal distribution as the sample size grows. Specifically, if $X_1, X_2, \ldots, X_n$ are random variables with mean μ and variance σ^2, the standardized sum:

$$Z_n = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sigma\sqrt{n}}$$

will converge to a standard normal distribution as n $\to \infty$

Considering our binary expansion of π where $X_i = 1$ with probability $p = 0.5$ and $X_i = 0$, we have $\mu = 0.5$ and $\sigma^2 = 0.25$. Dividing the sequence into segments of 500, we compute the means of these segments and scale them using the above formula to see if they resemble the standard normal distribution.

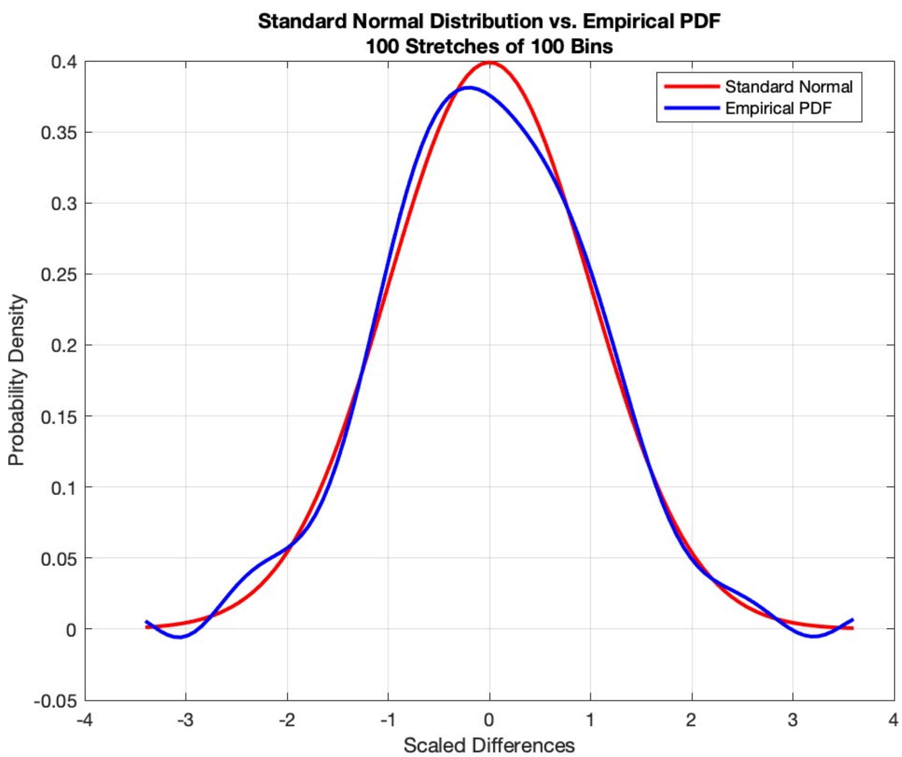

Figure 3: Distribution of scaled averages from segments of the binary sequence of π.

From the depicted histogram, we can observe a bell-shaped curve that closely resembles the standard normal distribution.

5 Conclusion

This analysis merges computational techniques with mathematical theory to demonstrate that the binary digits of π exhibit true randomness.