Assignment 1 - Randomized Algorithms

Dinu Eduard-Adiel

September 2024

1 Introduction

The Galton Box is a method or device to illustrate the central limit theorem.

The normal parameters for a specific experiment are as follows:

- n the number of levels or depth
- N the number of balls that will be participating in a given test

2 Simulator

There is no graphic interface for this assignment. The simulation consists of a simple vector representing the columns and a random choice for left or right movement.

A usual experiment consists of selecting a n and N and letting the simulator iterate through every ball.
This is the snippet of the simulator function:

```
def simulate_galton_board(n, N):
    counts = np.zeros(n + 1)
    for _ in range(N):
        position = 0
        for _ in range(n):
            move = np.random.choice([0, 1])
            position += move
        counts[position] += 1
    return counts
```

The true binomial and normal distributions are computed with the *scipy.stats* library: *binom.pmf* and *norm.pdf* and the required parameters.

3 Binomial and Normal distribution

Since at any moment t_i the movement choice is an independent event with a 2 possible answers: *left* or *right* the whole experiment of Galton Box can be seen as a binomial distribution and as a whole it can be seen as a Normal distribution.

On a superficial way of seeing the experiment, just by counting the number of paths from the center it can be very easy to deduce that the middle (the average or mean) will contain the most amount of balls and the number will start decreasing towards the margins (standard deviation).

Theorem 1 (Probability Mass Function of Binomial Distribution)**.** *Let X be a random variable representing the number of successes in n independent Bernoulli trials, each with a probability of success p. The probability mass function is given by:*

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

for $k = 0, 1, 2, \ldots, n$.

Theorem 2 (Probability Density Function of Normal Distribution)**.** *The probability density function of a normal distribution with mean μ and standard deviation σ is given by:*

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

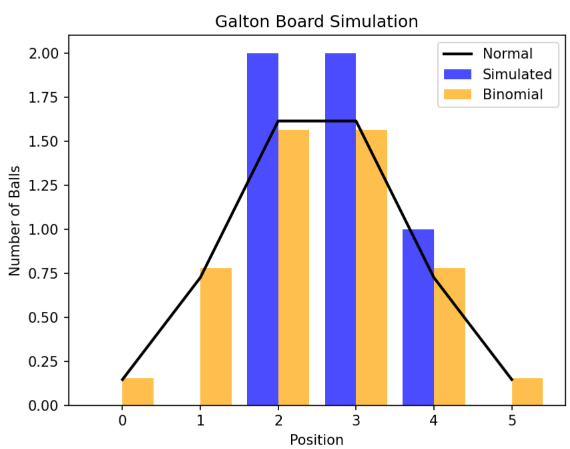
for all $x \in \mathbb{R}$.

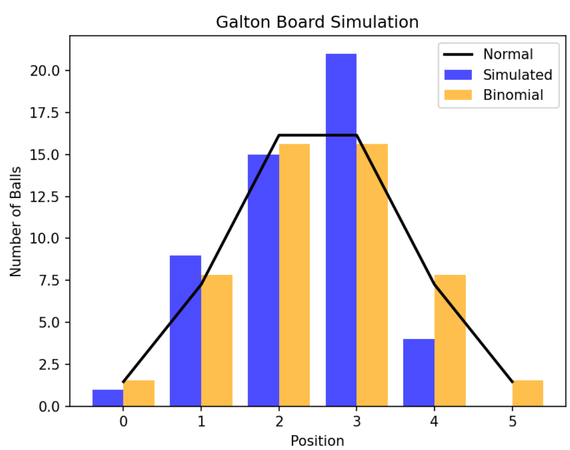
4 Experiments

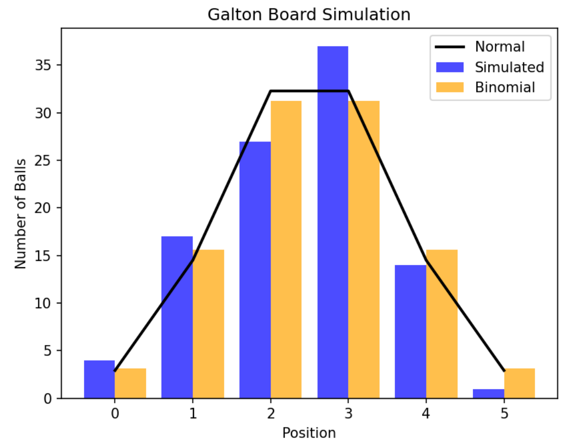
In this section, I will explain the parameters and the direction of my experiments.

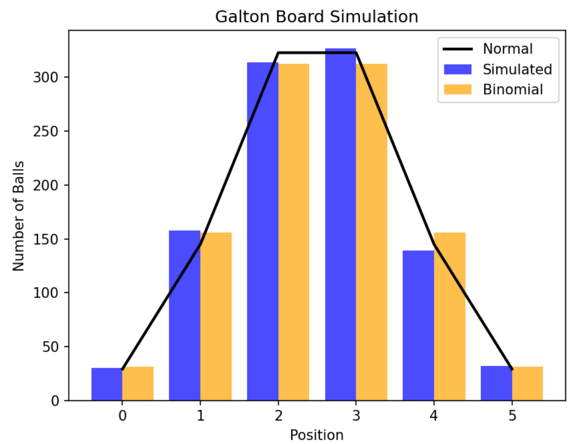
4.1 Number of Balls

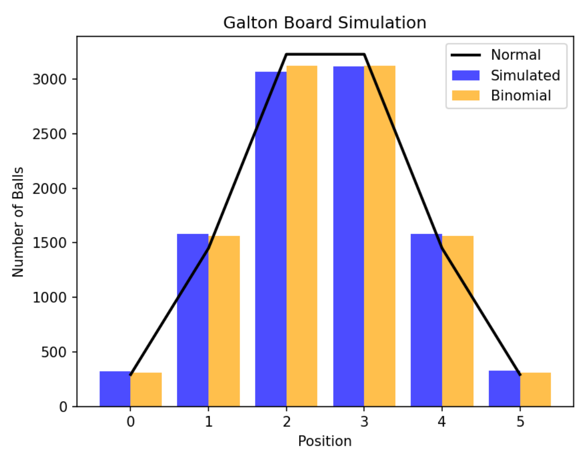
Here is a series of experiments based on the value of the number of balls. For consistency, the values will be taken from the set $S = \{5, 50, 100, 1000, 10000\}$. Below are shown the results of the simulated distribution, binomial and normal distribution.



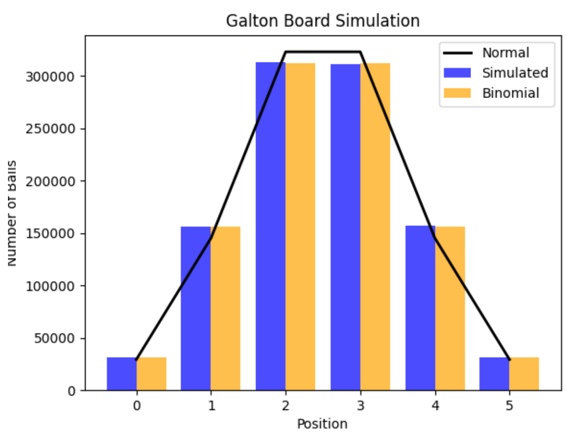








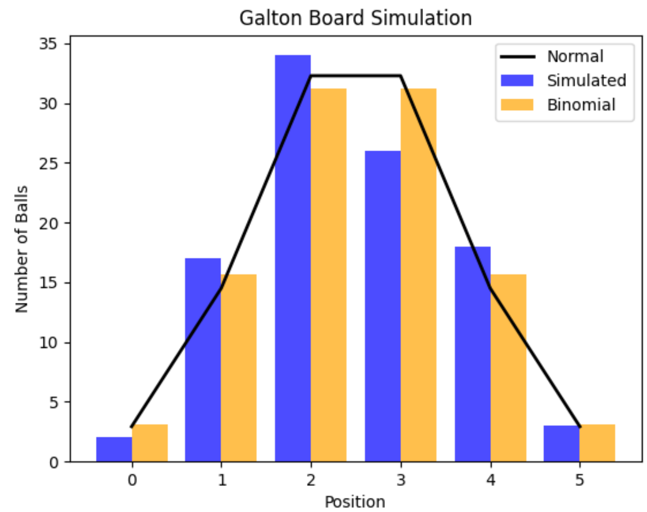
As seen in the previous graphs, when the number of balls is small the simulated distribution is irregular. By repeating the same experiments eventually, the simulated distribution will look like the binomial and normal one, but this in practice can take a long time. **But, as the number of balls grows larger and larger**, the simulation will begin to resemble the normal and binomial faster. In the next graph, N is very large and it can be seen how it drastically changes the distribution. As in the *frequentist approach*, as the number of iterations goes to infinity, the distribution becomes the real distribution.

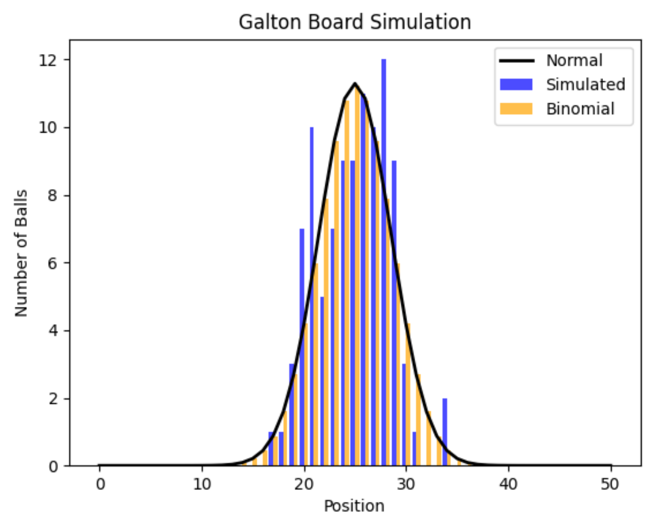


In this picture, it can be seen quite clear that the simulated is approximately the binomial distribution and if we convert the normal distribution to a discrete on, then it is almost the same.

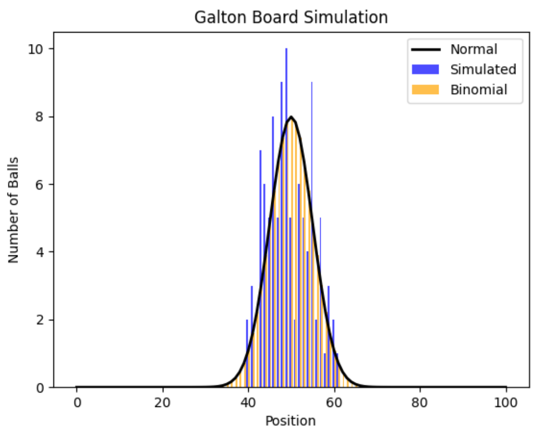
4.2 Number of levels

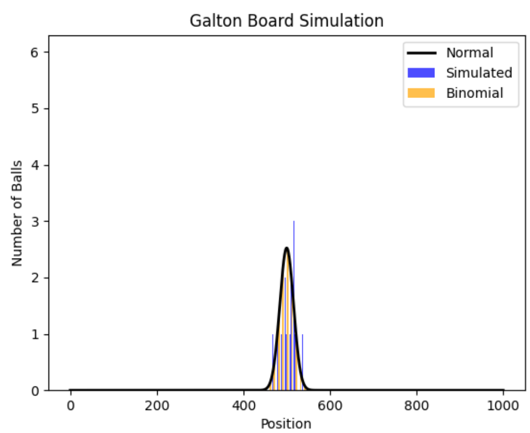
Here is a series of experiments based on the value of the number of levels. For consistency, the values will be taken from the set $n = \{5, 50, 100, 1000, 10000\}$ and the number of balls will be $N = 100$ as the starting point.

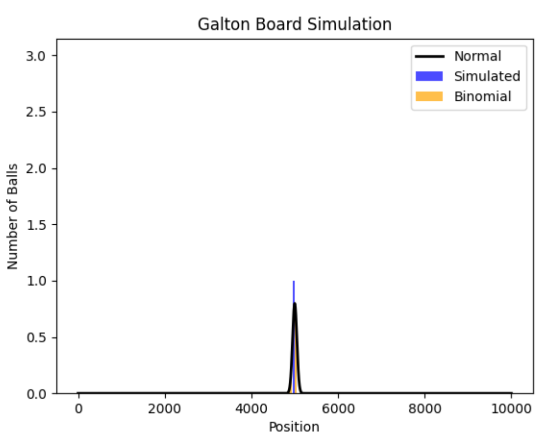




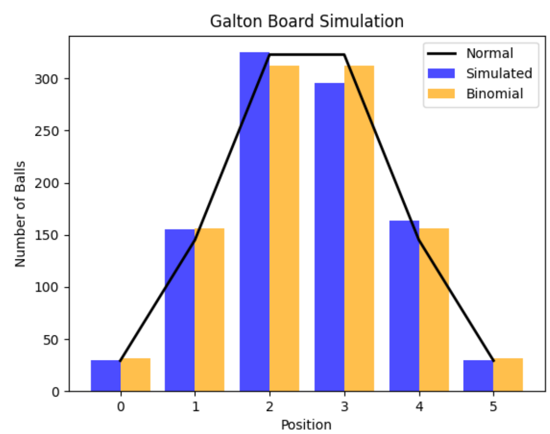
It can be seen that when n is smaller than N, the distribution starts to resemble the true distribtuion.

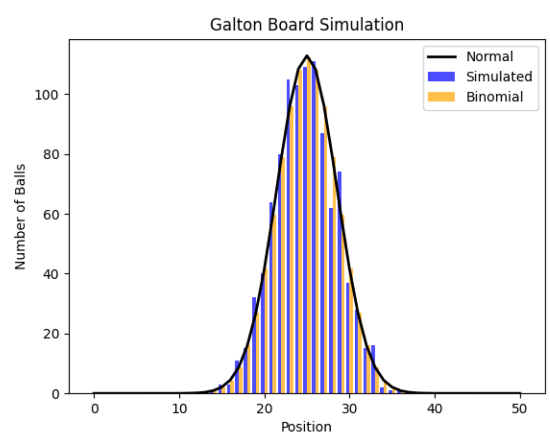


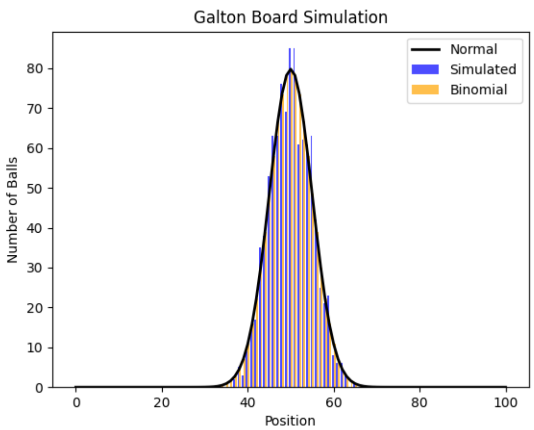


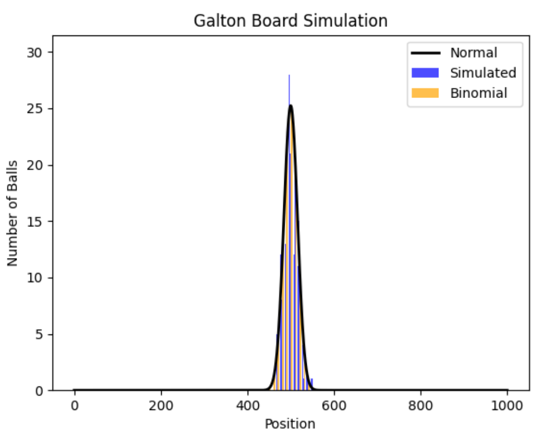


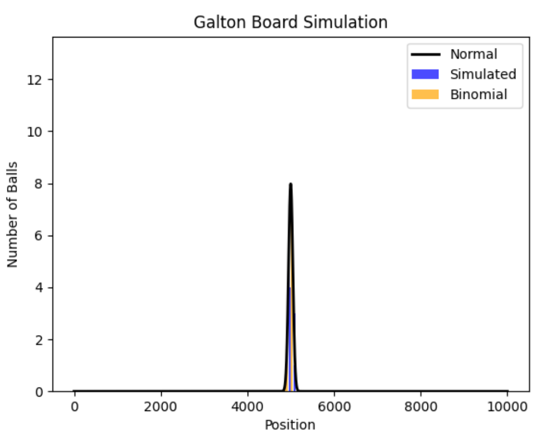
As seen in the previous subsection, N increasing means a bigger resemblance to the real distribution, however, here n is dependent to the number of experiments (or Bernoulli trials) for a larger and clearer view. Since N starts at 100 and from the third experiment, n is bigger than that, it means that there is a small chance that every level is filled so we need more iterations.



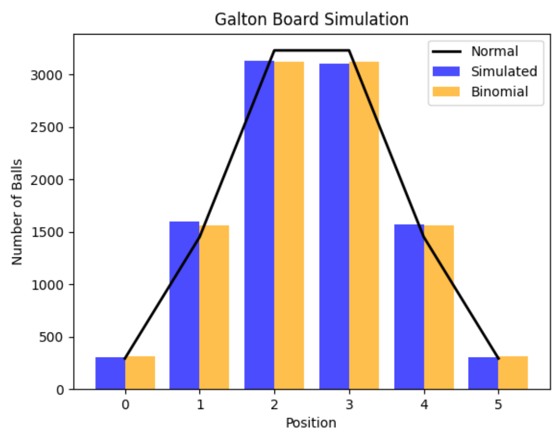


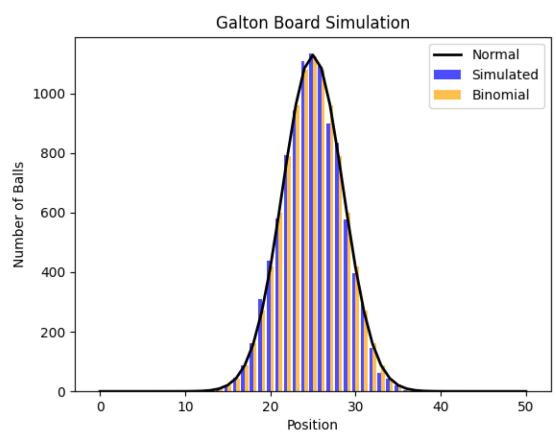


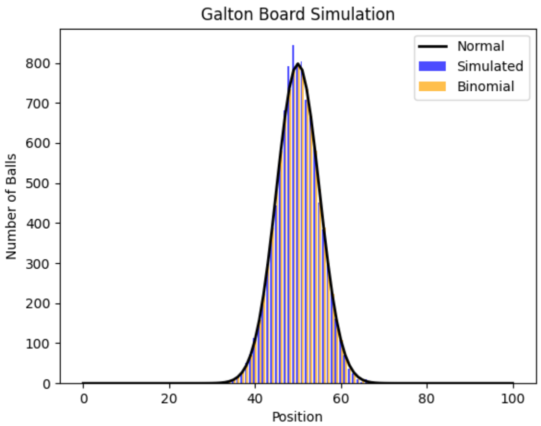


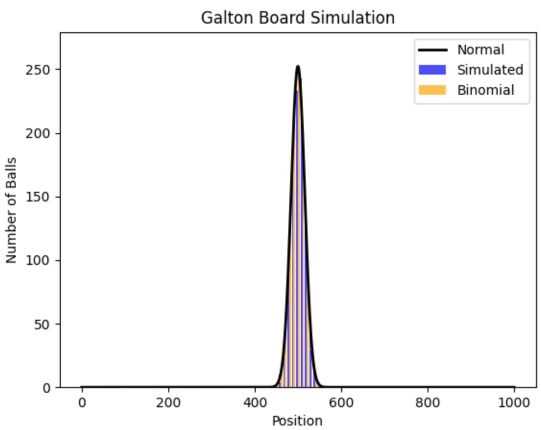


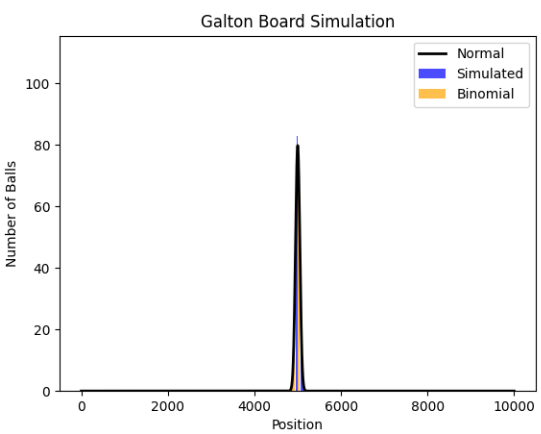
It becomes more clear of the resemblance, but here is an experiment with $N = 10000$ balls, for a more precise measurement. These results show much better the resemblance.



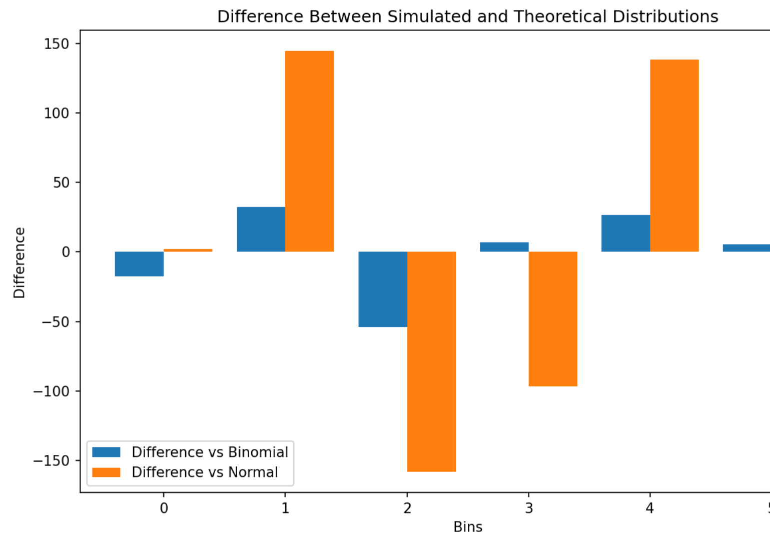


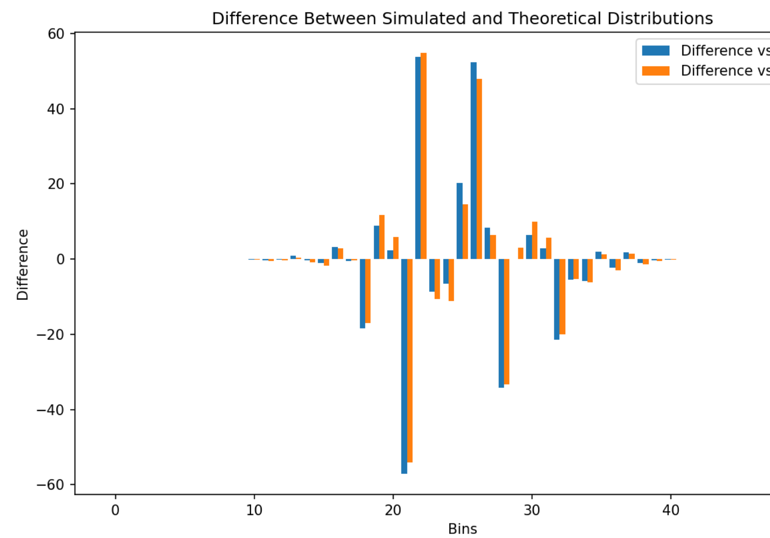


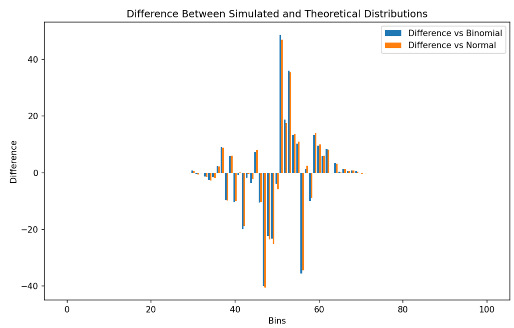


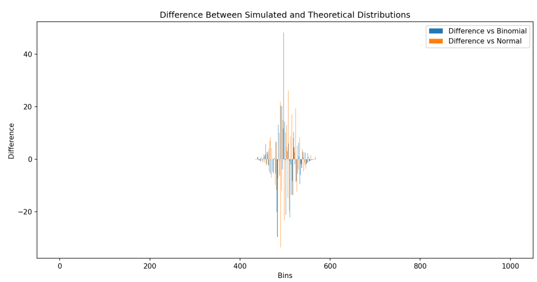


4.3 Difference between simulated and true distributions









As seen from the difference graphs, as seen n and N grow, the peripheries start to become identical while the main mean point and its vicinity become the center of attention, from the 10000 balls the difference is the level of 40s.

5 Conclusions

The Galton Box illustrates perfectly the central limit theorem. Its distribution accurately describes the Kolmogorov's approach to probability and it is a very useful tool for learning the properties of the binomial and normal distribution. As the number of balls grows the simulated distribution becomes clearer and even though the increase in the number of levels should help with the accuracy of the simulated distribution, it heavily relies on the increase in the number of balls.

6 Instructions for use

The main program is *galton.py* and it is run as a normal python program using the terminal command:

```
$ python galton.py
```

In order to change the parameters of the program, the 80 needs to be changed to the desired n and N.

The link to the repository is this:
`https://github.com/adieldinu/Randomized_Algorithms`

7 Bibliography

- `https://www.reddit.com/r/statistics/comments/193hn9w/q_what_is_so_interesting_about_the_galton_board/`
- `file:///C:/Users/adiel/Downloads/Dialnet-AScrewbiasedGaltonBoard-8444854.pdf`