Randomized Algorithms

Computer Science Department

Simulating Galton Board

Authors:
Alex Herrero
Lluna Clavera

Professor:
Conrado Martínez

September 28, 2025

Contents

List of Figures

List of Tables

1 Introduction

In this programming assignment, our goal is to simulate a Galton board in order to observe the probability distribution of the balls falling into cells, and to demonstrate that, as we increase the board size, the distribution gets closer to a normal distribution.

Given a Galton board of size n and N balls we will generate the simulation by generating, for each ball, n random bits. Every ball starts at position $(0,0)$, the ball in (i,j) will move to position $(i+1-b, j+b)$ where b is a random bit (a number that can be 0 or 1) generated previously. This process will be repeated for each generated random bit, until we arrive to a position of the form $(k, n-k)$ where $0 \leq k \leq n$.

All the code used for our experiments, as well as our results, can be found in the GitHub repository: `https://github.com/AleexHrB/RA-MIRI`

2 Probabilstic Analysis of the Galton board

Given an n-level Galton board, we let X_i be the random variable representing the number of balls that land in the i-th position (for $0 \leq i \leq n$). Since each hit in a cell is independent of the others, we can write

$$X_i = \sum_{j=0}^{n} Y_j,$$

where Y_j is a Bernoulli random variable that equals 1 if, at the j-th step of the Galton board, the random choice is to the right, and 0 otherwise.

As X_i is the sum of independent Bernoulli random variables, it follows that X_i can be modeled by a binomial distribution. Specifically,

$$X_i \sim \mathrm{Bin}(n, \frac{1}{2})$$

where n is the number of steps and the probability of moving right at each step is $p = \frac{1}{2}$.

We also know that the binomial distribution $Bin(n,p)$ can also be approximated by a normal distribution with mean $\mu = n \cdot p$ and $\sigma^2 = n \cdot p(1-p)$. In our case, that means that we can approximate the distribution of our balls with the normal $\mathcal{N}(\mu, \sigma^2) = \mathcal{N}(n/2, n/4)$, but the error of the approximation will decrease as we increase the size of the board (as we will check in a future chapter).

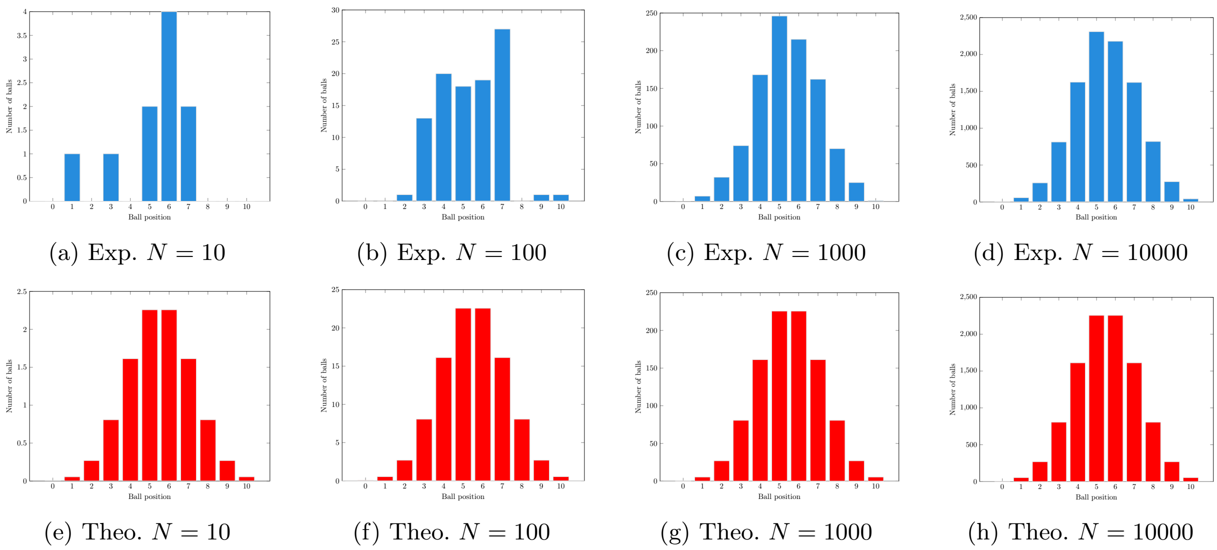
3 Experimentation

3.1 Galton Board

We first performed the experiment with a fixed Galton board of size $n = 11$ and varying ball counts $N \in \{10, 100, 1000, 10000\}$. The results can be seen in Figure 1, where we observe that as the number of trials increases with a fixed Galton board, the experimental results converge to the binomial distribution.

Next, we perform another experiment where we fix the number of trials while increasing the size of the Galton board (see Figure 2). In this case, the experimental results do not appear to converge to the theoretical distribution. However,it is reasonable to expect that they should still follow a binomial distribution (see Section 2). However, this is difficult to observe because the number of balls is smaller than the size of the board, which makes plotting the binomial distribution challenging.

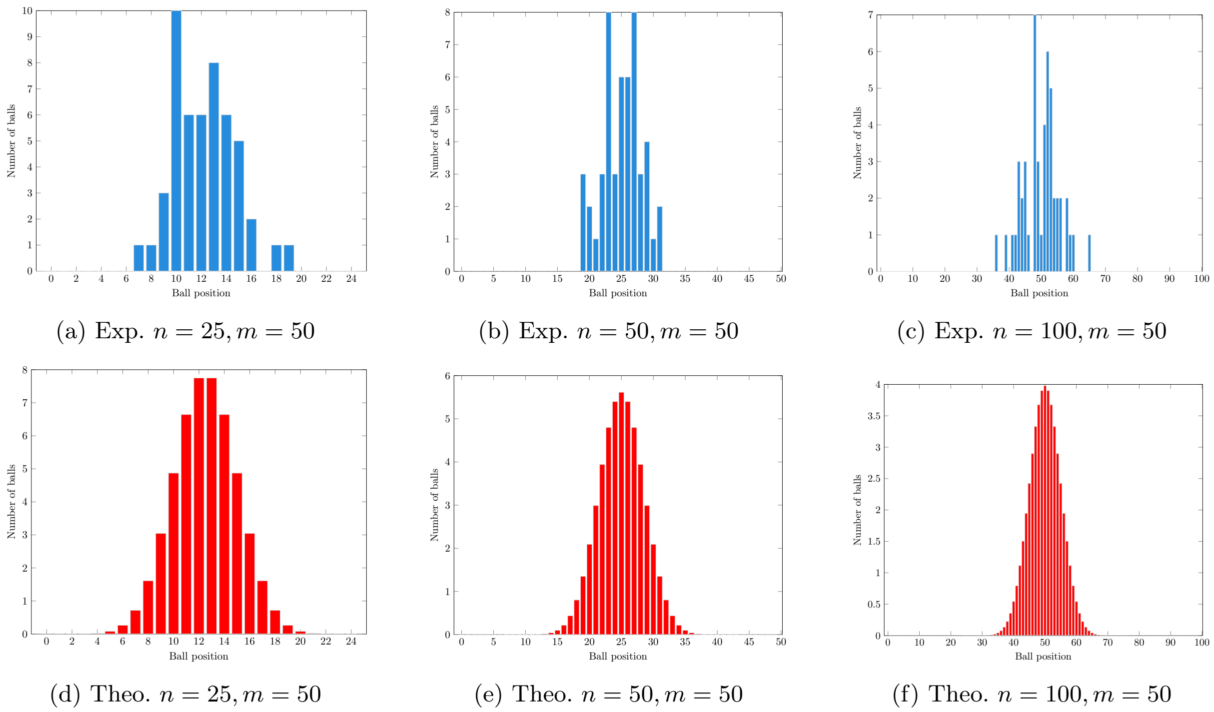
Finally, we also computed the maximum error and mean squared error (MSE) of our experiment and the theoretical results (given by the binomial distribution). We also computed the standard deviation at the cell where we got the maximum error (so we can compare them).



(a) Exp. $N = 10$ (b) Exp. $N = 100$ (c) Exp. $N = 1000$ (d) Exp. $N = 10000$

(e) Theo. $N = 10$ (f) Theo. $N = 100$ (g) Theo. $N = 1000$ (h) Theo. $N = 10000$

Figure 1: Comparison of experimental (top row) and theoretical (bottom row) Galton board results for $n = 11$ with different values of N.



(a) Exp. $n = 25, m = 50$ (b) Exp. $n = 50, m = 50$ (c) Exp. $n = 100, m = 50$

(d) Theo. $n = 25, m = 50$ (e) Theo. $n = 50, m = 50$ (f) Theo. $n = 100, m = 50$

Figure 2: Comparison of experimental (top row) and theoretical (bottom row) Galton board results with $m = 50$ for different values of n.

Experiment	Max Error	Standard Deviation of the Max	MSE	MSE/N²
$n = 11$; $N = 10$	$1.7441 (< 2\sigma)$	1.1626	0.6895	0.0069
$n = 11$; $N = 100$	$10.8867\ (= 4\sigma)$	2.7217	23.8732	0.0024
$n = 11$; $N = 1000$	$20.4141\ (< 2\sigma)$	13.2173	71.2164	$7.1216 \cdot 10^{-5}$
$n = 11$; $N = 10000$	$76.0547\ (> 2\sigma)$	36.7506	881.6384	$8.8305 \cdot 10^{-6}$
$n = 10$; $N = 50$	$2.9512\ (< 2\sigma)$	2.2515	1.4643	$6.0985 \cdot 10^{-4}$
$n = 25$; $N = 50$	$5.1292\ (> 2\sigma)$	2.3999	1.4049	$5.6194 \cdot 10^{-4}$
$n = 50$; $N = 50$	$3.2919\ (< 2\sigma)$	2.1943	0.7931	$3.1722 \cdot 10^{-4}$
$n = 100$; $N = 50$	$3.3295\ (< 2\sigma)$	1.8455	0.5329	$2.1314 \cdot 10^{-4}$

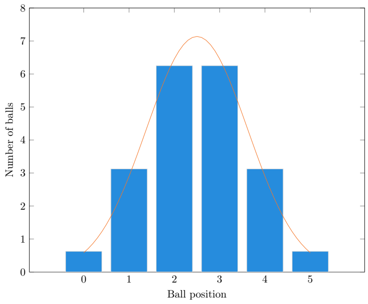
Table 1: Experimental Results

As we can observe in Table 1, the standard deviation as well as the MSE increases as we increase the number of balls. If we have a high number of balls we will expect a high number of balls to fall in each cell, so we can also expect to have a high deviation and, therefore, a high error. We can compensate for the number of balls by dividing the MSE over N^2. After dividing the error by the number of balls, we can observe that the error goes down. We can also observe that increasing the board size also decreases the error (even if we do not increase the number of balls), we theorize that this happens because we have more empty cells, which will have very small error and will bring the mean down.

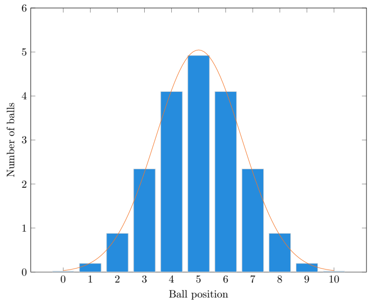
Additionally, with the obtained data we can compare the maximum error with the standard deviation. In table 1 we see that all maximum errors are higher than the standard deviation (which makes sense as we are picking the *maximum* error), but most are between one and two standard deviations, so they are not too far from the expected. However, we obtained two instances with an error between two and three standard deviations and one instance with an error equal to four standard deviations. Obtaining some instances with error greater than two standard deviations makes sense as we have done multiple experiments, but the instance at four standard deviations is really anomalous. We found this anomaly really odd, so we hypothesized that it could have been provoked by the pseudo-random number generation of C++, but after repeating that same experiment multiple times to check if there were more anomalies, we found none (The extra experiments can be found in the repository). Therefore we conclude that this anomaly is just a product of working with random variables.

3.2 Agreement between Binomial and Normal distributions

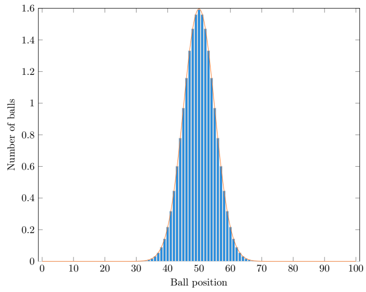
By the Central Limit Theorem we know that, for large value of n we can approximate the value of a binomial distribution X with probability p to $\mathcal{N}(np, np(1-p))$. In our case, since each attempt has a probability of success of 1/2, we now have $X \sim \mathcal{N}(n/2, n/4)$. As we see in Figure 3 we have better approximations of the binomial distribution with a normal one as we continue to increase n.



(a) Binomial vs Normal distribution with $n = 5$



(b) Binomial vs Normal distribution with $n = 10$



(c) Binomial vs Normal distribution with $n = 100$

Figure 3: Check between Binomial and Normal distribution with different values of n

Furthermore, Figure 4 illustrates the behavior of the Mean Squared Error (MSE) within this approximation context, demonstrating a rapid decrease in MSE as we increase the size n.

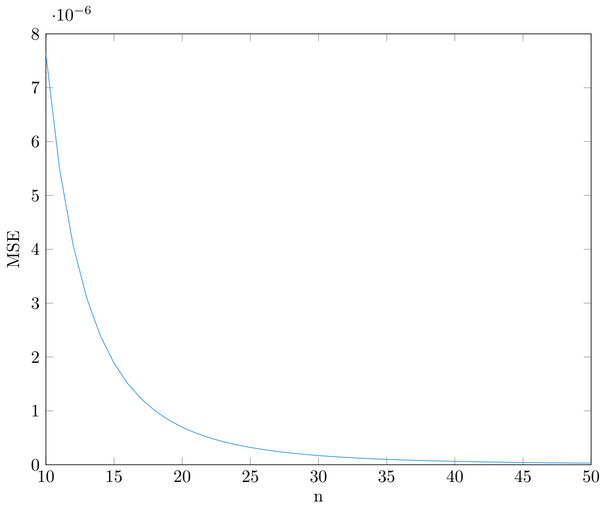


Figure 4: MSE between binomial and normal distribution