Randomized Algorithms (RA-MIRI): Assignment #1

Izan Beltran Ferreiro

September 2024

Contents

1 Introduction **2**

2 Probabilistic Analysis **3**

3 Simulation **5**

3.1 Randomness 5

3.2 Implementation 5

3.3 Results 5

3.3.1 Results on the distribution 5

3.3.2 Results on the mean squared error (MSE) 5

1 Introduction

A *Galton board* [4] (or *Galton box*) is a device used to ilustrate the central limit theorem. It consists of a vertical board with interleaved rows of pegs. A large number of small balls falls from the top center, bouncing left or right when they hit the pegs. The bottom of the board collects balls into bins that will follow the binomial distribution (see Figure 1).

Figure 1: A Galton board

2 Probabilistic Analysis

Notation. We'll use H to denote the number of rows or height of the Galton board and N to denote the number of balls. We enumerate the bins at the bottom of the Galton board from 0 to H.

Consider a ball falling from the Galton board. Initially, it is at horizontally centered on the board, but it moves H times to either left or right, each with probability 1/2. We can assign a random variable $R_j^i\,(1 \leq j \leq H, 1 \leq i \leq N)$ to each random bounce, indicating whether the i-th ball moves to the right on the j-th bounce.

Definition 2.1. Given $p \in [0,1]$, the Bernoulli distribution $\mathrm{Bernoulli}(p)$ is defined such that for any random variable $X \sim \mathrm{Bernoulli}(p)$,

$$\Pr[X = 0] = 1 - p, \quad \Pr[X = 1] = p$$

We define R_j^i such that $R_j^i \sim \mathrm{Bernoulli}(1/2)$. Now, we can assign a random variable P_i to each ball, indicating its final position. Therefore,

$$P_i := R_1^i + \cdots + R_H^i, \quad 1 \leq i \leq N \tag{1}$$

Proposition 2.1. *Let* $X \sim \mathrm{Bernoulli}(1/2)$. *Then,*

1. $\mathrm{E}[X] = p$
2. $\mathrm{Var}[X] = p(1-p)$

Proof. By definition of the expected value,

$$\mathrm{E}[X] = 0 \cdot \Pr[X = 0] + 1 \cdot \Pr[X = 1] = p$$

now, using the definition of variance,

$$\begin{aligned}\mathrm{Var}[X] = \mathrm{E}[(X - \mathrm{E}[X])^2] = \mathrm{E}[(X-p)^2] = (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p = \\ = p^2(1-p) + (1-p)^2 p = p[p(1-p) + (1-p)^2] = \\ = p(1-p)[1 + (1-p)] = p(1-p)\end{aligned}$$

□

Definition 2.2. Given $n \in \mathbb{Z}$ and $p \in [0,1]$, the Binomial distribution $\mathrm{Bin}(n,p)$ is defined such that for any random variable $X \sim \mathrm{Bin}(n,p)$,

$$\Pr[X = k] = \binom{n}{k} p^k (1-p)^k =: p_{n,k}, \quad k \in \mathbb{Z}, 0 \leq k \leq n$$

Proposition 2.2. *The Binomial distribution is equivalent to the sum of* n *independent Bernoulli distributions. Formally,*

$$\mathrm{Bin}(n,p) \sim \underbrace{\mathrm{Bernoulli}(p) + \mathrm{Bernoulli}(p) + \cdots + \mathrm{Bernoulli}(p)}_{n} \tag{2}$$

Proof. Let $X \sim \mathrm{Bin}(n,p)$ and $Y = Z_1 + \cdots + Z_n$, where $Z_i \sim \mathrm{Bernoulli}(p)$ $(1 \leq i \leq n)$ and all of them are independent. By definition of X, given an integer $0 \leq k \leq n$, we have $\Pr[X = k] = p_{n,k}$. Let's compute $\Pr[Y = k]$. We need k of the Z_i to be equal to 1, and the remaining ones equal to 0. Such assignment to the variables will have probability $p^k(1-p)^k$, independently of the order of the results. From combinatorics we know there are $\binom{n}{k}$ such assignments. Therefore, $\Pr[Y = k] = \binom{n}{k}p^k(1-p)^k = p_{n,k}$, which implies $X \sim Y$. □

Corollary 2.0.1. *Since P_i is the sum of n Bernoulli distributions of probability 1/2, the bins of the Galton board follow a Binomial distribution of probability 1/2. Formally, $P_i \sim \mathrm{Bin}(n, 1/2)$.*

Definition 2.3. Let C_k be the random variable counting the numbers of balls in the k-th bin.

Proposition 2.3. *If $X \sim \mathrm{Bin}(n,p)$, the following holds*

1. $\mathrm{E}[X] = np$
2. $\mathrm{Var}[X] = np(1-p)$

Proof. Since X is distributed as the sum of n independent Bernoulli distributions of probability p, by linearity of expectation we have that $\mathrm{E}[X] = n\,\mathrm{E}[\mathrm{Bernoulli}(p)] = np$. Now, by linearity of variance of the sum of independent variables, we have $\mathrm{Var}[X] = n\,\mathrm{Var}[\mathrm{Bernoulli}(p)] = np(1-p)$ □

Proposition 2.4. *Let C_k be the random variable counting the number of balls that end up in the k-th bin of the Galton board. Then $C_k \sim \mathrm{Bin}(N, p_{n,k})$.*

Proof. Each ball is independent of each other and has probability $p_{n,k}$ of landing at the k-th bin. Therefore it follows a Binomial distribution. □

Corollary 2.0.2. *$\mathrm{E}[C_k] = p_{n,k}N$, $\mathrm{Var}[C_k] = p_{n,k}(1-p_{n,k})N$ and C_k has standard deviation $\sqrt{p_{n,k}(1-p_{n,k})N}$.*

Definition 2.4. The Normal (or Gaussian) distribution $\mathcal{N}(\mu, \sigma^2)$ is defined such that its probability density function is

$$f(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Theorem 2.1. (Central Limit Theorem, [3]) *Let $X_1, \ldots, X_n$ be independent identically distributed random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$ $(1 \leq i \leq n)$. Let $\overline{X}_n := (X_1 + \cdots + X_n)/n$. Then, as $n \to \infty$, $\sqrt{n}(\overline{X}_n - \mu)/\sigma^2$ converges in distribution to $\mathcal{N}(0,1)$.*

This gives us a theoretical justification for having $\overline{X}_n$ distributed like $\mathcal{N}(\mu, \sigma^2/\sqrt{n})$. Since the number of balls in the k-th bin is $C_k \sim \mathrm{Bin}(N, p_{n,k})$, we have that C_k is distributed approximately like $\mathcal{N}(\mu', \sigma'^2)$, where $\mu' = p_{n,k}N$ and $\sigma'^2 = p_{n,k}(1-p_{n,k})\sqrt{N}$. For the positions of balls, $P_i \sim \mathrm{Bin}(H, 1/2)$, we have that the mean is distributed approximately like $\mathcal{N}(H/2, H/4)$.

3 Simulation

3.1 Randomness

Turing machines are deterministic and require of an external source of randomness in order to execute randomized algorithms. There are two main ways to solve this problem:

- Use pseudorandom generators. They behave closely to true random sequences for most purposes.
- Use an actual device which allows the computer to extract true random data.

The rest of this section applies to both cases indistinctively, since on practise results will be similar for our experiment. However, different kind of experiments may require the actual usage of randomness in order to give accurate realistic results.

3.2 Implementation

The implementation [1] allows the user to choose using pseudorandomness or true randomness (implemented for a Linux device with a properly working `/dev/random` [2]).

In order to simulate a ball throw from height H, we just use H coin flips to determine how many bounces are to the right. This is a Binomial distribution with probability $1/2$ and H trials. This uses $\Theta(H)$ calls to the random generator.

To simulate the full Galton board, we execute N ball throws of height H, with a total number of $\Theta(NH)$ calls to the random generator.

3.3 Results

For the sake of reproducibility, the results presented here use a fixed seed which can be found on the implementation source code. Execution with true random generators yields similar results.

3.3.1 Results on the distribution

We conducted the experiment for $H \in \{10, 100\}$ and $N \in \{10, 10^2, 10^3, 10^5\}$ and plotted the balls distributions over the bins, along with the expected (Binomial) distribution and the Normal distribution. Figures 3 and 4 show that the results converge in distribution to a normal distribution as N/H gets big.

3.3.2 Results on the mean squared error (MSE)

We further investigated the mean squared error with the experimental results and both the binomial distribution and the normal distribution, and also plotted them (Figure 2) in a logarithmic scale for multiple combinations of (H, N).

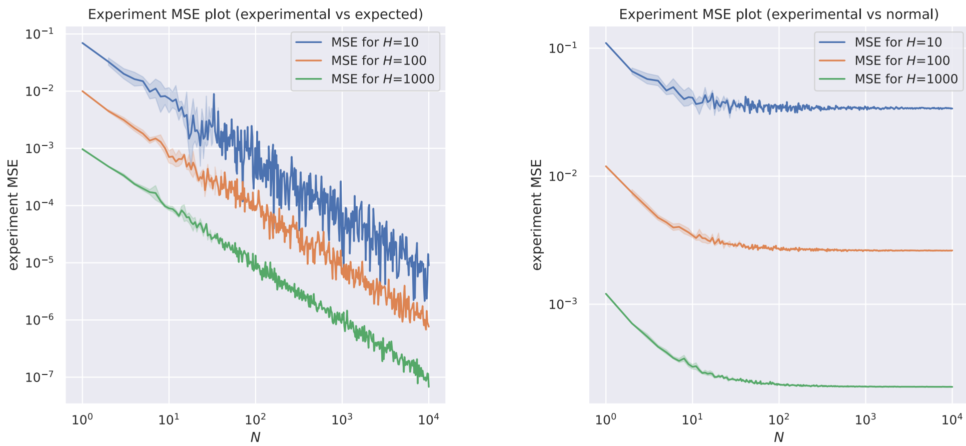

Figure 2: MSE plots

On the MSE with the binomial distribution we observe a log-linear pattern that suggests the error is about $\Theta(\frac{1}{NH})$. This implies that increasing both parameters gives a linear improvement on the difference between the obtained results and the expected result. Another interesting observation is that variance in experimental results decreases only when H increases.

On the other hand, the MSE with the normal distribution show an initial decrease with both parameters but rapidly converges to a fixed value with N and only improves linearly with H. This is expected since the parameter which makes the binomial approach a normal distribution is H, the number of bounces of the balls.

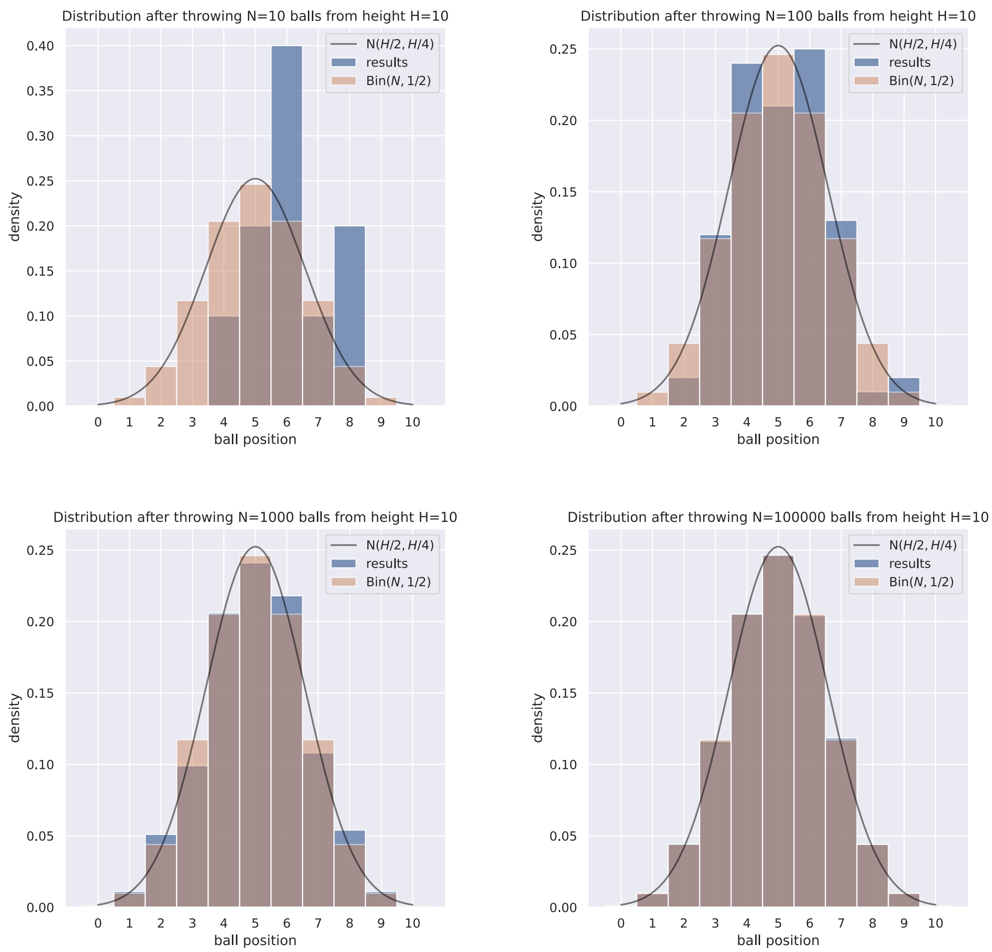

Figure 3: Experiment results for H=10

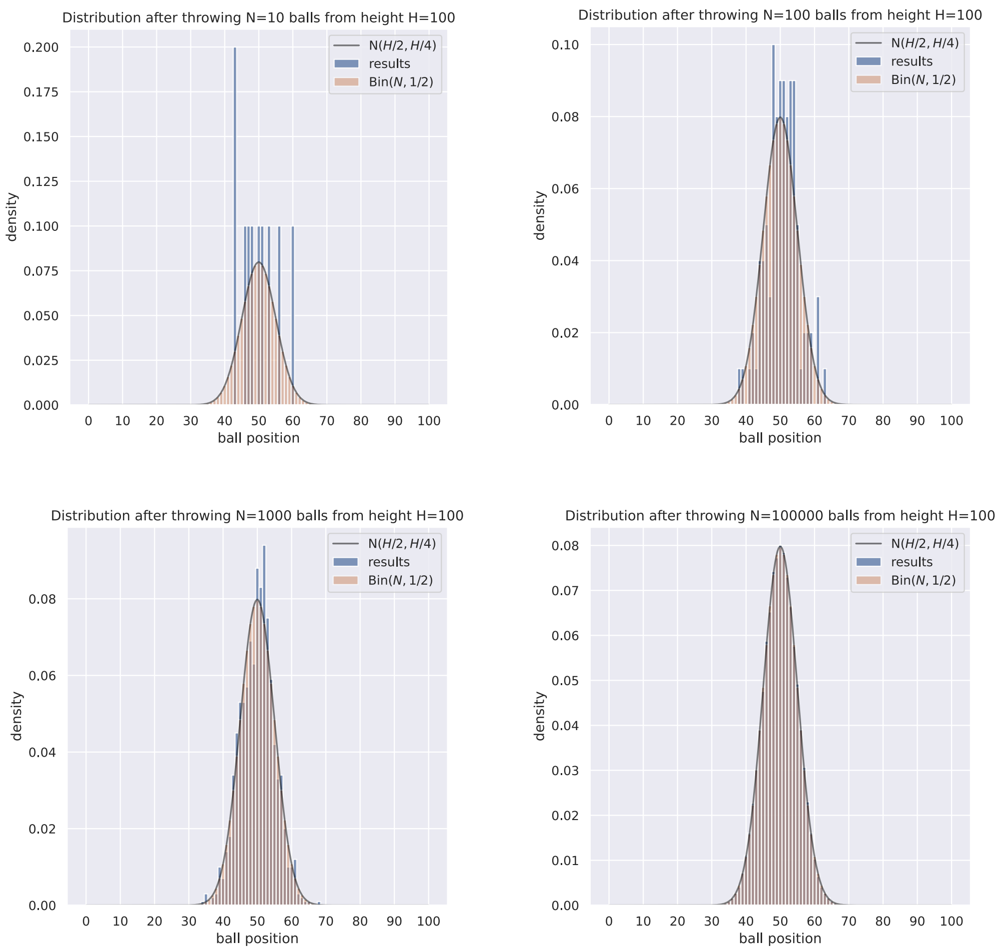

Figure 4: Experiment results for H=100

References

[1] Izan Beltran. *Implementation source code*. URL: https://github.com/izanbf1803/galtonbox.

[2] wikipedia.org. */dev/random*. URL: https://en.wikipedia.org/wiki//dev/random.

[3] wikipedia.org. *Central limit theorem article*. URL: https://en.wikipedia.org/wiki/Central_limit_theorem.

[4] wikipedia.org. *Galton board*. URL: https://en.wikipedia.org/wiki/Galton_board.