Randomized Algorithms (RA-MIRI): Assignment #1

1 Statement of the assignment

In this programming assignment you will have to write a program to simulate a *Galton board*, a.k.a. *Galton box*, quincunx or bean machine, a device used to illustrate the central limit theorem, in particular, that with sufficently large sample the binomial distribution is approximated by the normal distribution. The device consists of a vertical board with interleaved rows of pegs. A large number of small balls fails from the top, their path to the bottom of the board bouncing left and right when the balls hit the pegs. The bottom of the board collects balls that will follow the binomial distribution (see Figure 1).

Figure 1: A Galton board

We can simulate the Galton board with a virtual triangle matrix. Balls "drop" from cell $(0, 0)$ up to n levels. At each step of the simulation, we have a ball in some cell (i, j); with probability $1/2$ the ball moves to the "left", that is, cell $(i + 1, j)$, or with probability $1/2$ to the "right", to cell $(i, j + 1)$. After n steps the ball will be at some cell $(i, n - i)$ (because at each of the n steps we increase by 1 either the row or the column; the sum of row and column must be n).

The probability $p_{i,n}$ that a ball starting at $(0, 0)$ ends at cell $(i, n-i)$ is given by a binomial (why?). If we throw N balls, we expect to get $N \cdot p_{i,n}$ balls landing at cell $(i, n - i)$. The standar deviation is $\sqrt{N p_{i,n} q_{i,n}}$, where $q_{i,n} = 1 - p_{i,n}$.

Write a program to carry out the experiment and study the match between the experimental data and the predictions of probability theory. Study the effect of large boards (increase the value of n) and of a large number of experiments = a larger number N of balls. Check also the agreement between the binomial distribution and the normal distribution $\mathcal{N}(\mu, \sigma^2)$. Notice that $p_{i,n} \to 0$ as n grows, then the approximation between the binomial $\text{Bin}(n, 1/2)$ and the normal distribution $\mathcal{N}(n/2, n/4)$ is much better, and that's more visible when we draw a larger number N of samples of the $\text{Bin}(n, 1/2)$ distribution.

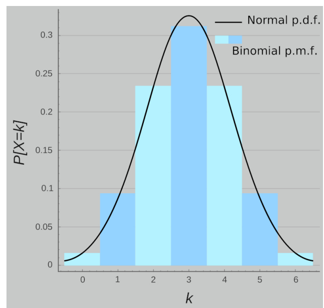

Figure 2: Binomial vs normal dsitributions.

Use graphical plots to illustrate the outcomes of your experiments. Quantifying (and plotting/tabulating) the error between empirical results vs theoretical predictions is also important, and how that error evolves as we vary the relevant parameters, namely, n and N. For example, do not content yourself producing drawings similar to that of Figure 4; instead, compute the mean quadratic error between the pdf of the normal distribution (the solid black line) and the probaility mass function of the binomial distribution (heights of the blue rectangles) at the integer points $k = 0$, $k = 1$, ...

2 Instructions to deliver your work

Submit your report in PDF using the FIB-Racó. The deadline for submission is October 14th, 2024 (8:00). The submitted file should be called

`YourLastName_YourFirstName-galtonbox.pdf`

The report must include also a link to a repository (in Github or GitLab, for example) or shared folder which contains all the source files of your program(s) and a `README` file with instructions to compile and execute the program(s) to reproduce the experiments. Make sure that the link works. It is also a good idea to provide a `.zip` or `.tar` file, with all the necessary files to carry out the assignment, ready for downloading.

N.B. I encourage you to use LaTeX to prepare your report. For the plots you can use any of the multiple packages that LaTeX has (in particular, the bundle

TikZ+PGF) or use independent software such as matplotlib and then include the images/PDF plots thus generated into your document.