Galton Box Simulation

Michele Bonomi

October 2025

Repository link: `https://github.com/ohhmeco/GaltonBox-simulation`

A Galton board, a.k.a. Galton box is 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. In that box, balls are dropped from a single funnel at the top, centered over the pins. We see the Galton Box as a triangular matrix in which at each step of simulation the ball moves to the left or to the right with probability 1/2. As a ball falls, it encounters pins at each level and at every pin, the ball has an equal probability (typically p=0.5) of bouncing left or right. The ball's path is in fact a random walk that ends in one of the vertical bins at the bottom and the final bin index corresponds to the net number of right turns the ball made.

Why the probability that a ball starting at cell (0,0) ends at cell (i, n-i) is given by a binomial?

We say that if the box has n rows of pins, a ball's path can be modeled as a sequence of indipendent Bernoulli trials, so the probability distribution of the number of times a ball goes right is a Binomial Distribution with parameters n standing for rows and p, standing for having gone right.

1 Simulation and Analysis

The analysis of the Galton Box simulation confirms the predictions of the Central Limit Theorem and shows the influence of both the number of levels (n) and the sample size (N). As demonstrated in Figure 1 (small n, small N), the empirical distribution shows significant sampling variance, although it is clearly centered around the mean $\mu = n/2$. This high variability is indeed expected with small sample sizes. When the sample size is increased significantly while keeping n small (e.g., $n = 10$, $N = 10000$ in Figure 3), the Binomial distribution remains the most precise theoretical model out of the two, with the Normal approximation exhibiting a higher Mean Squared Error (MSE). This confirms that the approximation Bin $\approx$ N is less reliable for small n.

When the number of levels n is large (e.g., $n = 100$, as shown in Figure 2), the Binomial distribution itself becomes symmetric and smooth.Here, the MSE values for the Binomial and Normal distributions converge, confirming that the Normal distribution is a good and computationally simpler approximation for the Binomial. The lowest relative sampling error is achieved with the largest board and largest number of balls ($n = 100$, $N = 100000$, Figure 4), where the empirical data shows the closest fit to both theoretical curves.

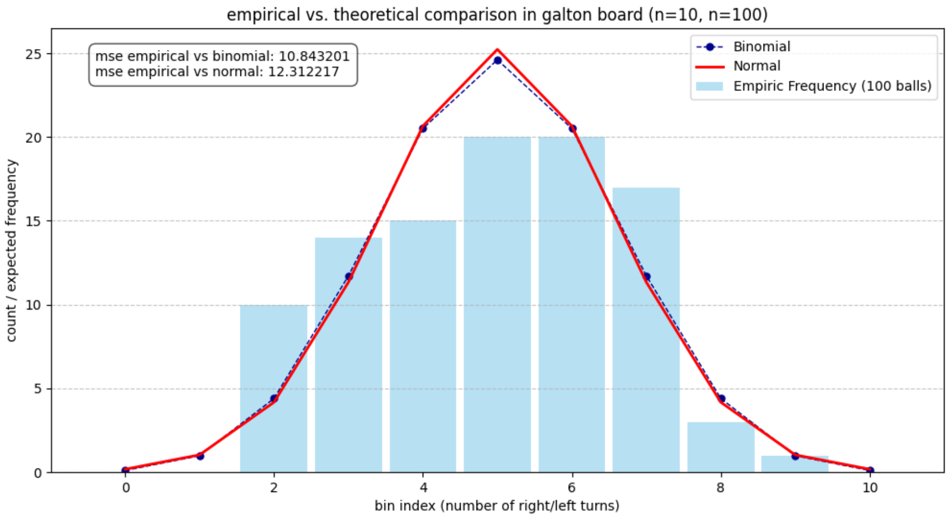

Figure 1: Experimental vs. Theoretical Distribution for small parameters ($n = 10$, $N = 100$). Note the high variance due to small N.

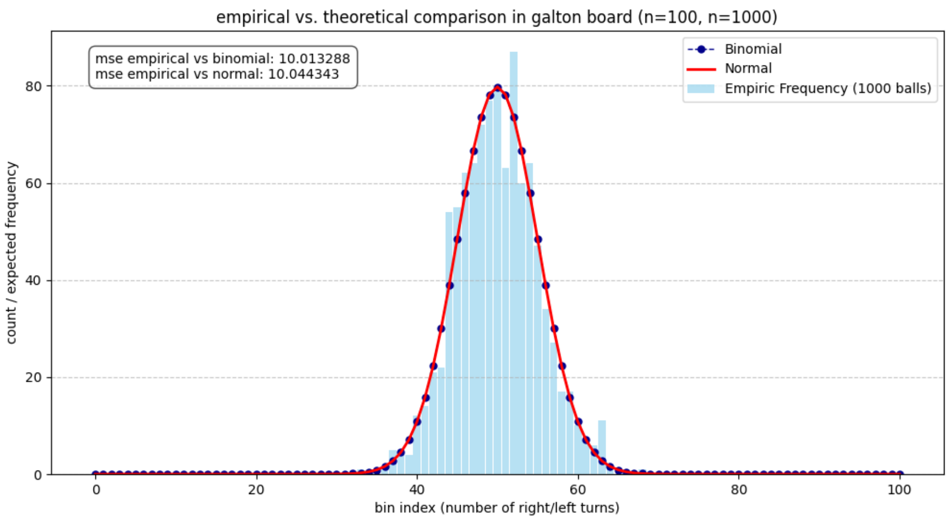

Figure 2: Distribution for large base ($n = 100$) but moderate sample size ($N = 1000$). The theoretical approximation is highly accurate.

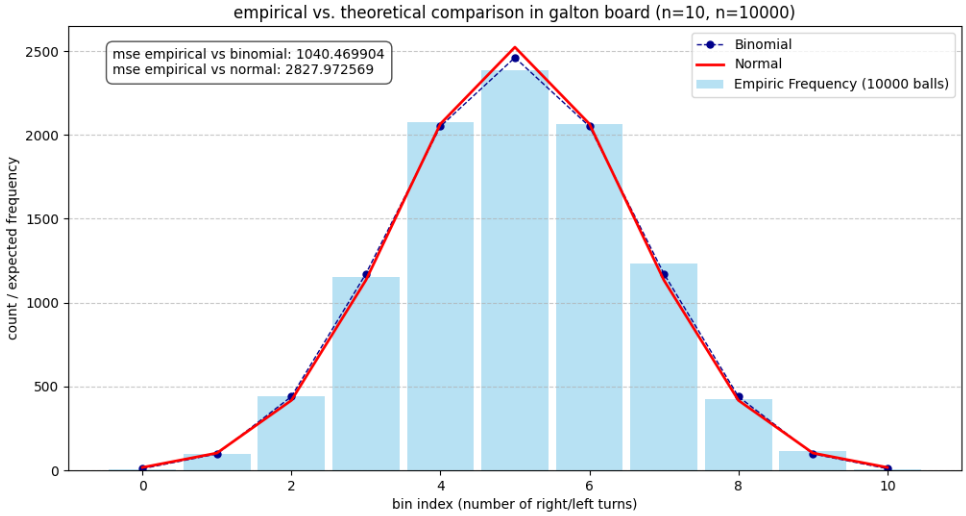

Figure 3: Distribution for small base ($n = 10$) with large sample ($N = 10000$). The Normal approximation is less precise than the Binomial, especially on the tails.

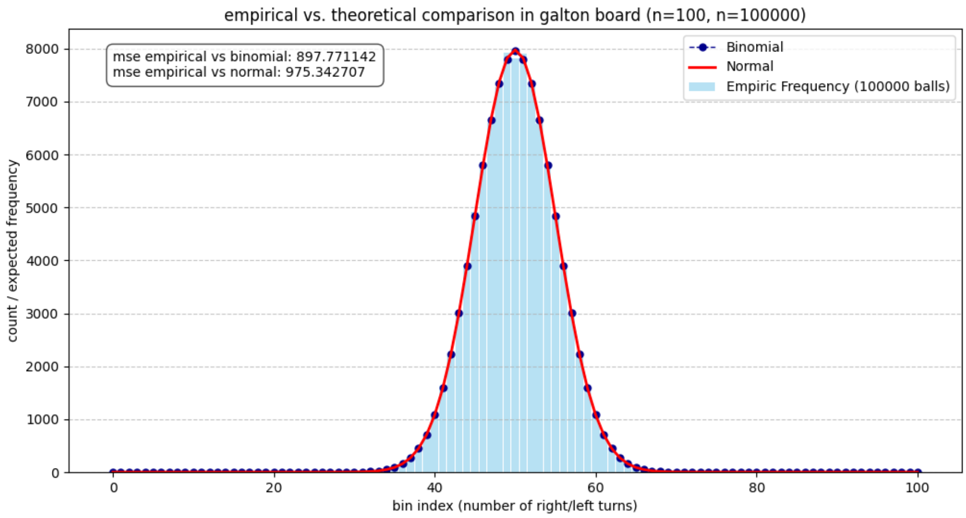

Figure 4: Distribution for large parameters ($n = 100$, $N = 100000$). Empirical data closely matches both theoretical distributions.