Stat. Phys. II: Numerical experiments on the binomial distribution

Emmmanuel Bossy

17 septembre 2025

I Definitions

One consider the random variable

$$M = \sum_{i=1}^{N} \epsilon_i$$

where each ϵ_i follows a Bernouilli law $P(\epsilon_i = +1) = p$ and $P(\epsilon_i = -1) = 1 - p$. M is thus a random variable which follows a binomial distribution with parameters N et p.

N^+ is the random variable defined as the number of positive ϵ_i in a given random sample, and N^- is the random variable defined as the number of negative ϵ_i in a given random sample. One thus has

$$N^+ = \frac{1}{2}(M + N)$$
$$M = 2N^+ - N$$

II General objectives

Theoretical predictions available for the binomial law includes value for $< M >$, $\mathrm{var}(M)$, $\mathrm{std}(M)$, $< N^+ >$, $\mathrm{var}(N^+)$, $\mathrm{std}(N^+)$, as well as the probability distribution $P(M = k)$ itself.

The objectives of these introductory sessions is to perform and analyze "numerical" experiments, to compare the results to these well known theoretical predictions. The process would be exactly the same for random variables with unknown probability distribution, you will just test it on a well-known distribution.

A significant amount of your time will likely be dedicated on using Python, which is part of the objectives...

III Specific tasks

For *all* graphical results, superimpose theoretical predictions and check for consistency based on compatibility with error bars. All final results should be compiled into a short .pdf report. Part of the job is to get familiar with very practical aspects such as getting figures with sufficient resolution, chosing relevant "experimental parameters" to compare "experiments" to theory, checking the consistency between "measurements" and "theory", in a situation where the theory is the right one.

1. Recall the relevant theoretical predictions.

2. Plot $< M >_p$ as a function of N, with error bars. $< M >_q$ is your ESTIMATE of $< M >$ obtained from q realizations. Choose a value of q that is relevant to illustrate how error bars depend on N.

3. Same for $\mathrm{var_p}(M)$ and $\mathrm{std_p}(M)$

4. Same for $< N+ >_p$, $\mathrm{var_p}(N+)$ and $\mathrm{std_p}(N+)$

5. Plot $\mathrm{std_p}(N+)/ < N+ >_p$ as a function of N, with error bars.

6. Plot histograms of M, for $N = 10$, $N = 100$, $N = 1000$, $N = 10000$, with error bars on each bin. Choose the size of the histogram bin in the most relevant way, as well as the range of the bins. Superimpose predictions from the theory, taking into account the size of the bins.

In the tasks above, estimating the error bars is a crucial aspect, as agreement with theory and experiments will depend on compatibility with error bars. It is however a non-trivial task....