Experimental verification of 10-coin, N=100 binomial distribution

Laboratory Notebook #1

Lyle Kenneth Geraldez (2019-11336)

theory-and-objectives:~# Distribution functions summarize the behavior of a given random variable X. Phenomenological modelling uses descriptive and/or inferential statistics to describe and predict system behavior. Often, central position <x>, variations σ, and correlation r concisely describe raw data characteristics; such quantities can be directly inferred from a given distribution function.

A boolean-valued random variable follows a binomial distribution function corresponding to C!/(C-x_j)!x_j! microstates. As an example, tossing C=10 coins follows a (normalized) binomial distribution function of form f_j=10!/2^10(10!-x_j)x_j! for a given x_j microstate (total heads). Mean value can be extracted as <x>= Σf_jx_j and standard deviation (s.d.) as σ=sqrt(Σ(x_j-<x>)^2f_j). To test these, we aim to:

! To experimentally verify binomial distribution pattern by physically-tossed set of ten coins and simulated pseudo-random number generator representing the setup,
! To calculate the descriptive statistics measure of central tendencies: (i) mean and (ii) mode through experiments and simulation and assessing deviation from theory at N = 100,
! To calculate the descriptive statistics measure of variability: (i) the standard deviation, through experiments and simulation and assessing deviation from theory at N = 100

setup-sketch~# A set of ten coins were simultaneously flipped N=100 times done by physically tossing similar coins and generating random numbers from Python

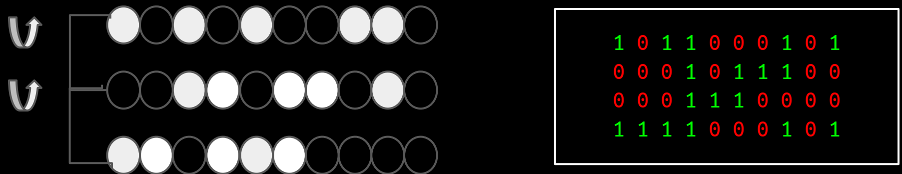

Figure 1. A set of ten binary integers following binomial distribution was extracted from physical results of tossing coins (left) and a seeded RNG (right).

The respective physical setup can be seen here: https://bit.ly/3Cl97Uh. For the physical coin toss, ten coins were tossed at the same time, allowed to flip freely and settle at an equilibrium state. Coins were individually checked and the number of head-faced coins were tallied. Data were stored in a csv file for processing. Similarly, we simulate an equivalent boolean phenomena in Python by generating an 10-element array of 1s and 0s to represent heads and tails, respectively. For replication purposes, RNGs were seeded. The process were, each, repeated N=100 times for both processes. For each microstate representing total heads per flip, a histogram was plotted for experimental and simulation data with bin = 10. Normalizing the histogram into a normalized distribution function, mean, mode, and standard deviations were calculated. Theoretical percent deviations were calculated by juxtaposing with the following theoretical binomial distribution:

Macro	Micro	f_j	Macro	Micro	f_j	Macro	Micro	f_j
0	1	9.8e-4	4	210	2.1e-1	8	45	4.4e-2
1	10	9.8e-3	5	252	2.5e-1	9	10	9.8e-3
2	45	4.4e-2	6	210	2.1e-1	10	1	9.8e-4
3	120	1.2e-1	7	120	1.2e-1			

results~# The study experimentally verifies the occurrence of a binomial distribution in a set of ten coins flipped N=100 times.

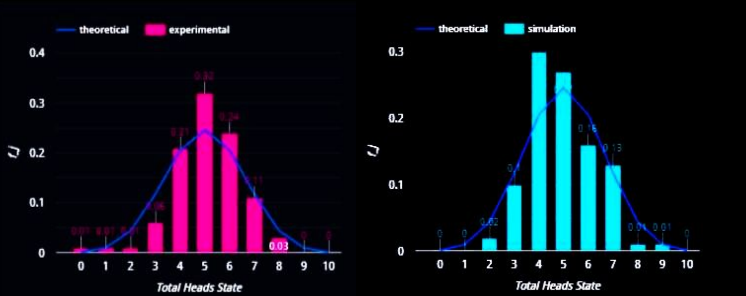

Figure 2. Bar charts of normalized empirical distribution functions for experimental (left) and simulation (right) with plotted theoretical binomial distribution

For theoretical data,
! <x>=5.00
! mode=5
! σ=1.58

For experimental data,
! <x>=5.10
! mode=5
! σ=1.34

For simulated data,
! <x>=4.93
! mode=4
! σ=1.37

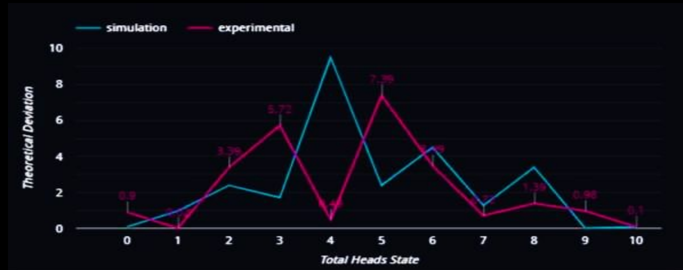

Figure 3. Absolute deviation of experimental and simulation data from theoretical distribution

For replication,

Raw data: https://bit.ly/3ST7GTQ
Data processing: https://bit.ly/3TaGLCP
Plots: https://bit.ly/3q2bVvc
Code repository: https://bit.ly/3vsqFDq

discussion~# Qualitatively, empirical data from both experimental and simulation data follows the characteristic trend of a binomial distribution function (blue curve interpolated from the previously shown table). Quantitatively, percent deviations were calculated from (|theory-empirical|/theory)*100 from each states and averaged across all data giving us 2.24% for experimental data and 2.40% for simulated data.

From the calculated f_j values. Experimental data gives us 2.0% deviation from theory while simulation data gives us 1.40% deviation from theory. The experimental mode matches theory while simulation mode shifts into x_j=4. S.d. for experimental has 13.1% deviation and simulation has 15.1% deviation from theory.

Observe an apparent out-of-phase nature of deviation between simulation and experiments. Although larger N is needed for a more convergent behavior to theory, simulated results at N=100 gives our experimental results a 99.87% degree of accuracy from a sampled (pseudo)ideal 50-50 coin toss.

conclusions~# The study verifies the following:

! Experimental and simulation results for 10-coin system flipped N=100 times show qualitative characteristics of binomial distribution; 97.76% accurate for experimental and 97.60% accurate for simulations.
! Mean value was successfully extracted from resulting distribution; 98.0% accurate for experimental and 98.60% accurate for simulations. Experimental mode matches theoretical mode.
! Standard deviation was successfully extracted from resulting distribution; 86.9% accurate for experimental and 84.9% for simulations.