Exercise 1: Probability Theory and Statistics

Versuchsdatum: 27.05.2025

TECHNISCHE UNIVERSITÄT DARMSTADT

Datenbasierte Modellierung
Maschinelles Lernen
ETIT

1 Solutions

1.1 Problem 1

1.1.1 Item (a)

Problem 1 (10 pts)

(a) A fair coin is tossed n times. Simulate a Bernoulli random variable with success probability $p = 0.5$. At each iteration k, compute a sample average of all k sampled elements. Afterwards, produce a plot of the average vs iterations. According to the law of large numbers, the sample average approaches the mathematical expectation, as $n \to \infty$. Take $n = 10^3$. **(5 pts)**

Figure 1.1: Item 1a

According to the Weak Law of Large Numbers (WLLN) the sample average tends to approach the expected value of the random variable. Since the random variable follows a Bernoulli distribution the expected values is equal to the probability of success p. It is possible to observe from Figure 1.2 that as the number of experiments grows, the mean of the sample tends to 0.5, which is equal to the probability of sucess used in the simulation.

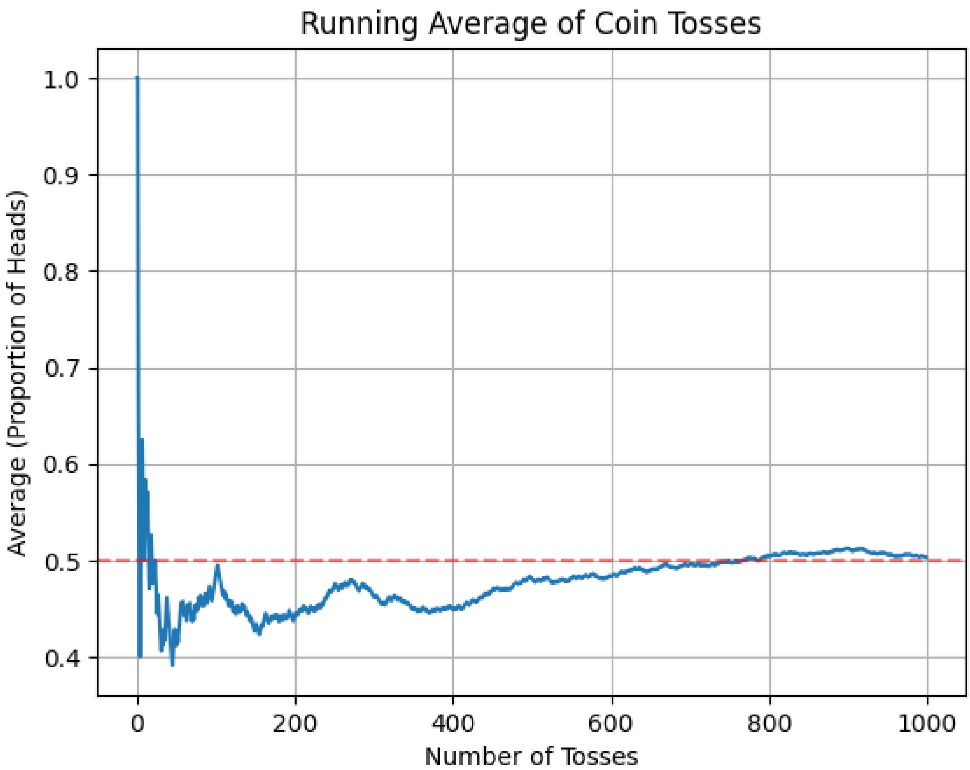
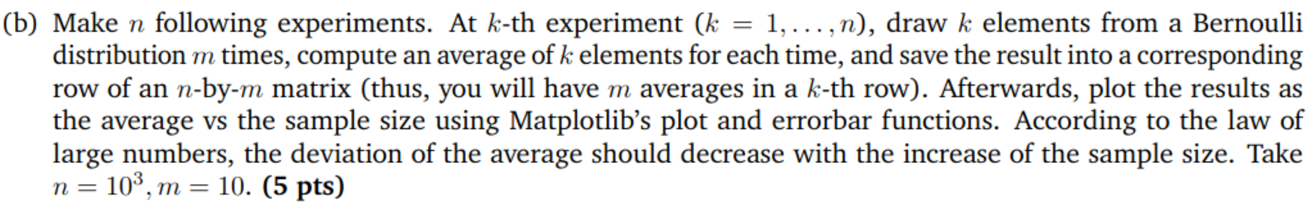


Figure 1.2: Running Average of Coin Tosses. n=1000. Equal Probabilities (p=0.5)

1.1.2 Item (b)



(b) Make n following experiments. At k-th experiment ($k = 1, \ldots, n$), draw k elements from a Bernoulli distribution m times, compute an average of k elements for each time, and save the result into a corresponding row of an n-by-m matrix (thus, you will have m averages in a k-th row). Afterwards, plot the results as the average vs the sample size using Matplotlib's plot and errorbar functions. According to the law of large numbers, the deviation of the average should decrease with the increase of the sample size. Take $n = 10^3, m = 10$. **(5 pts)**

Figure 1.3: Item 1b

In Figure 1.4 it can be seen that not just the sample mean tends to the expexted values of the distribuition, which is equal to the probability of success for a Bernoulli distribution, but also that the error bar limits tend to reduce as the number of samples increases, indicating that the deviation of the sample mean is being reduced as the sample size increases, as proposed by the WLLN.

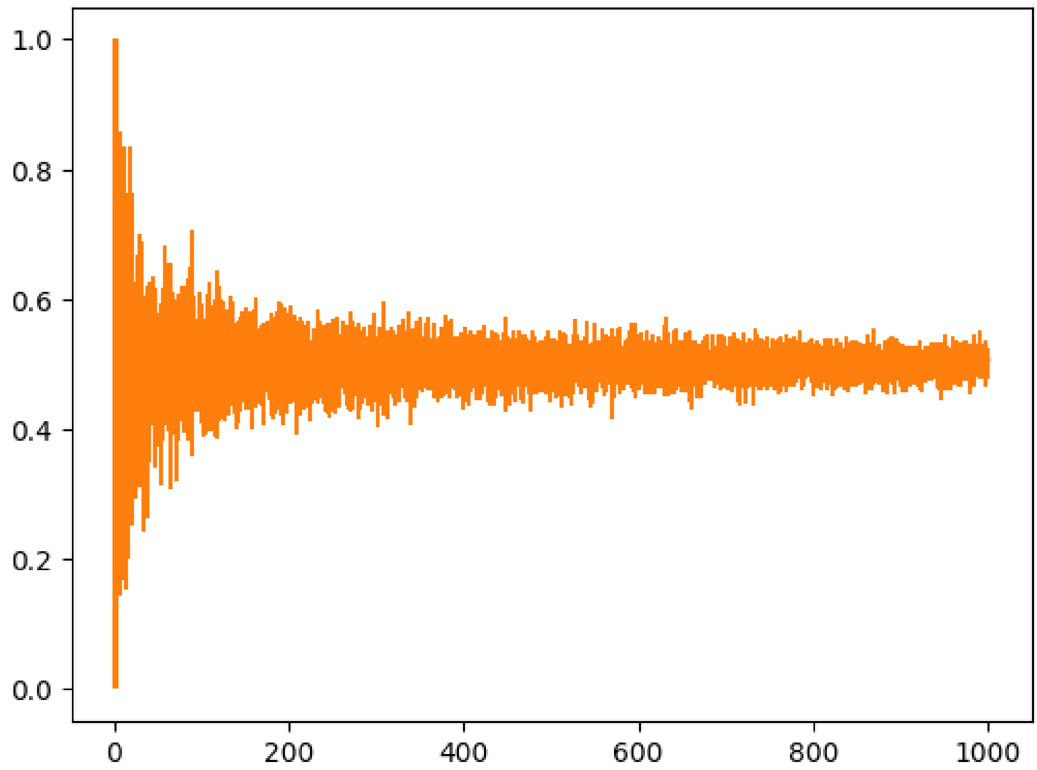


Figure 1.4: Deviation of Average for different values of sampling size.

(a) The sum of n independent Bernoulli random variables with success probability p has a binomial distribution with parameters n, p. Conduct the following procedure m times. Use the program from Problem 1 a., compute the number of successes for n tosses, and save this number into a corresponding row of an m-by-1 vector. After the procedure is done, draw a binomially-distributed random variable m times with parameters n, p, and save it into another m-by-1 vector. For both vectors, plot the resulting probability distributions with the help of a histogram and compute the mean-square error of the difference. Use the following parameter values: $n = 10^3, p = 0.3, m = \{10^3, 10^4, 10^5\}$. **(5 pts)**

Figure 1.5: Item 2a

From Figure 1.4, it is possible to observe that as the number of procedures m grows, the histogram of the number of successes for each experiment (left) tends to approximate the simulated binomial random variable (right). This behavior can be ascertained by the Table 1.1.

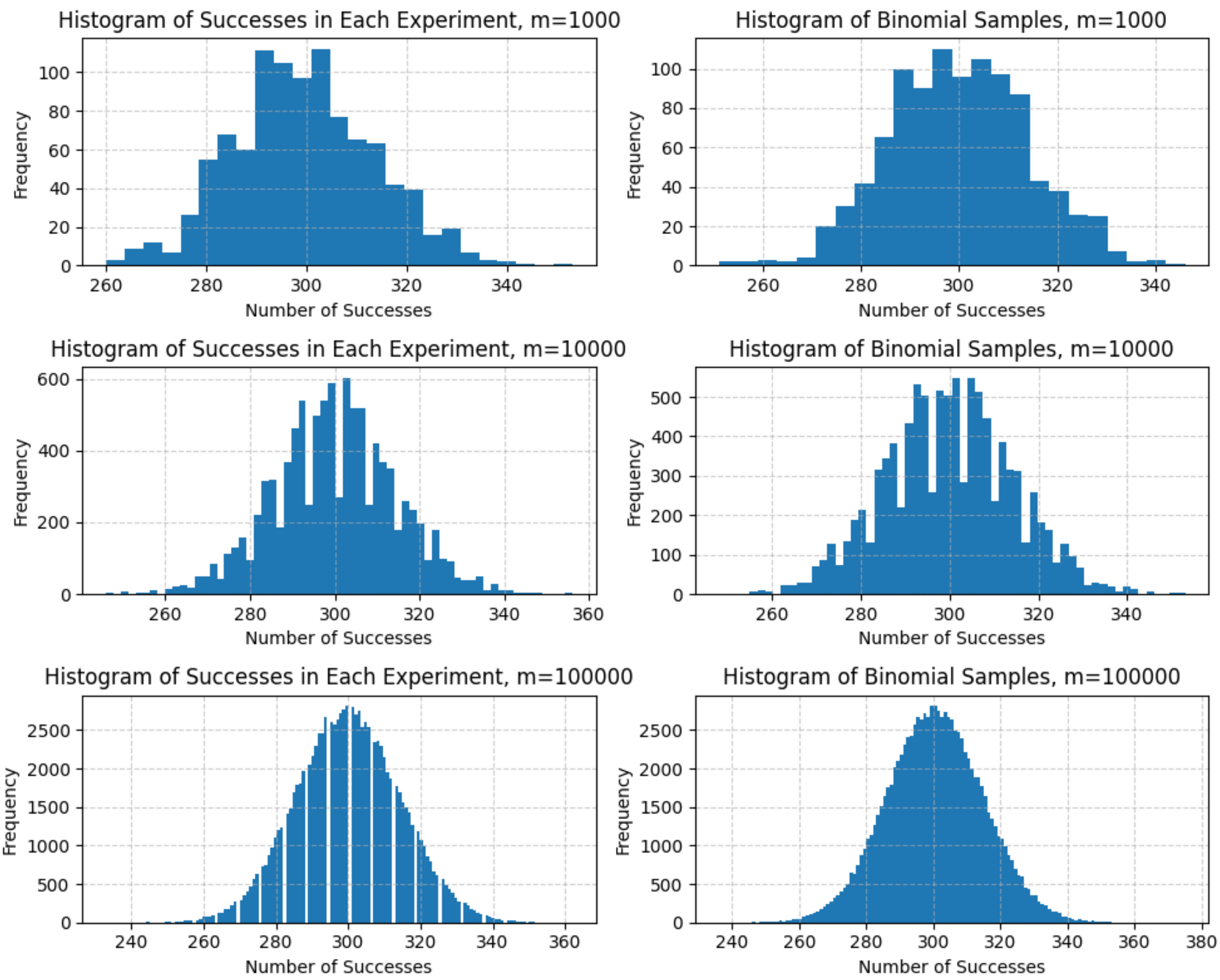


Figure 1.6: Histograms for different m values

m	1000	10000	100000
MSE of the difference (binomial - successes)	421.371671	423.441980	418.838612

Table 1.1: Variation of the Mean Squared Error between the total number of success and simulation of Binomial Samples with increasing values of experiments (m)

1.2.1 Item (b)

(b) According to the Poisson limit theorem, as $n \to \infty$ and and $p \to 0$, the Binomial(n, p) distribution approaches Poisson(np) distribution. Assume $n = 10^3, p = \{10^{-1}, 10^{-2}, 10^{-3}\}, m = 10^5$ and modify the previous program in order to compare Binomial and Poisson distributions. **(5 pts)**

Figure 1.7: Item 2b

As the probability of success p gets reduced for both distributions, the simulations of the binomial and Poisson tend to approximate. This can be seen not only in Figure 1.8, but also in the Table 1.2, in which the mean error and the Mean Squared Error tend to reduce as the value of p decreases.

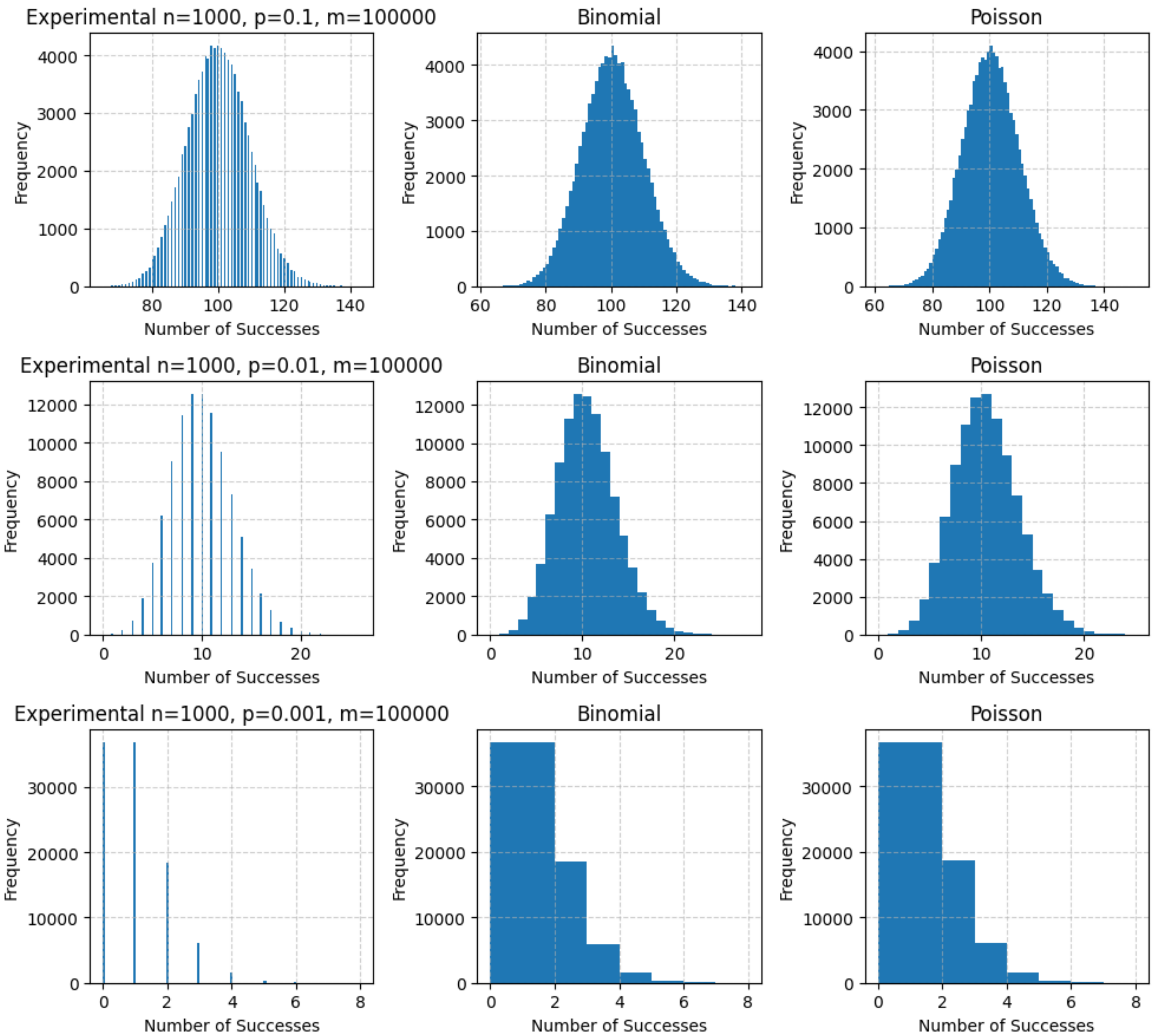


Figure 1.8: Each column show the results for different values of p

p	0.1	0.01	0.001
Mean Error with binomial	-0.00675	0.00233	-0.00172
MSE binomial	180.2968044375	19.875664571100003	1.9886970416
Mean Error with Poisson	0.01699	0.01691	0.00563
MSE Poisson	189.333721	19.672404	1.996018
Mean Error between Poisson and Binomial	0.01699	0.01691	0.00563
MSE between Poisson and Binomial	189.33372133989997	19.672404051900003	1.996018303100000

Table 1.2: Comparison of the Mean Squared Error and the Mean Error between the total number of success and the simulation of Poisson Samples with decreasing values of p. There are also comparisons between Binomial distribution in order to compare the behavior between the number of successes and the Binomial distribution