Practice 1

Central Limit Theorem

The objective of this practice is to calculate the probability distribution, average value and variance of throwing many dice multiple times and calculating the distribution of the outcomes.

In this problem, we will throw N_{dice} dice, N_{iter} times.

1.1 First Task

Use a random number generator to simulate dice throwing and calculate the resulting probability distribution p_i of obtaining each number from 1 to 6. Generate N_{iter} random numbers and calculate the mean μ

$$\mu = \langle x \rangle = \frac{1}{6} \sum_{i=1}^{6} i = 3.5$$

and the variance σ^2

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 = \frac{1}{6} \sum_{i=1}^{6} i^2 - (3.5)^2 \simeq 2.9167$$

or a standard deviation of $\sigma = \sqrt{2.9167} \simeq 1.7078$. Plot the resulting probability distribution function and compare it with a uniform distribution function.

1.1.1 Results

For this first part, we rolled the die $N_{\text{iter}} = 100.000$ times and obtained a relative frequency of

$$x_i = \{0.16662, 0.16777, 0.16495, 0.16826, 0.16717, 0.16523\}$$

for which we can estimate the expected outcome or the mean as $(\sum_{i=1}^{6} i \cdot x_i)/6 \simeq 3.49728$, which is very close to the exact result of 3.5. We can do the same to estimate the standard deviation, and obtain a value of 1.70616. Again very similar to the exact standard deviation. In figure 1.1, we can see the probability distribution of our random dice rolls compared with the uniform distribution, which is what we would expect to obtain after many attempts.

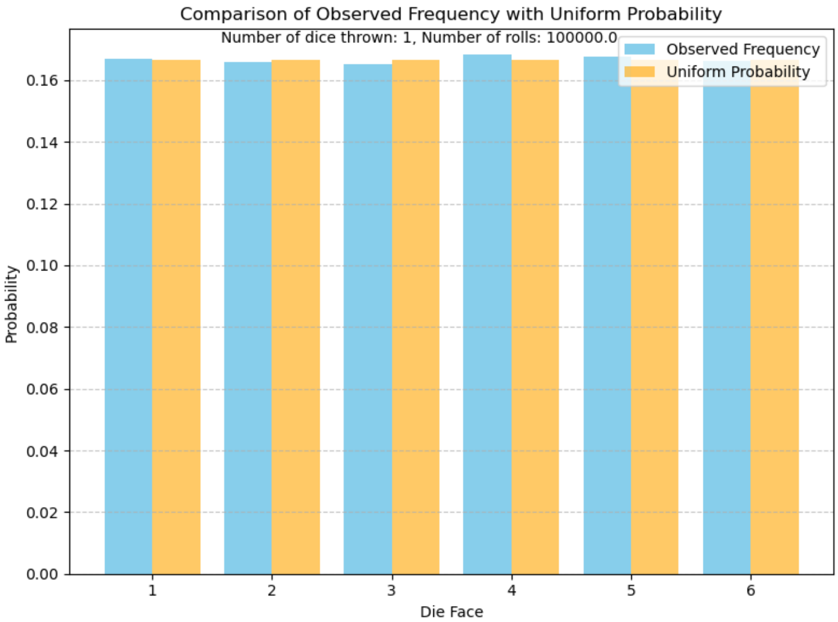

Figure 1.1: Observed frequency versus uniform probability

1.2 Second Task

Now, instead of throwing a single die, assume that one random event consists in throwing two dice ($N_{\text{dice}} = 2$) and calculating the average value

$$x_i = \frac{\text{rand}(6) + \text{rand}(6)}{2}$$

Calculate the probability distribution of the outcome $x = (1, 1.5, 2, 2.5, \ldots, 5.5, 6)$ and show it on a figure.

1.2.1 Results

In this case, while the mean value of the outcome did not change ($\mu = 3.5$), the exact standard deviation did. And comparing our estimations with these exact values we see that

- Probability distribution: [0.02718 0.05491 0.08417 0.11108 0.13821 0.16579 0.13892 0.11148 0.08431 0.05573 0.02822].
- $\mu = 3.5$ vs estimated mean: 3.5052 $\rightarrow$ relative error of 0.149%.
- $\sigma = 1.2076$ vs estimated std: 1.2082 $\rightarrow$ relative error of 0.047%.

Notice that in this case the exact probability distribution is the convolution of two uniform probability distributions, that is, a triangular distribution (as seen in fig.1.2).

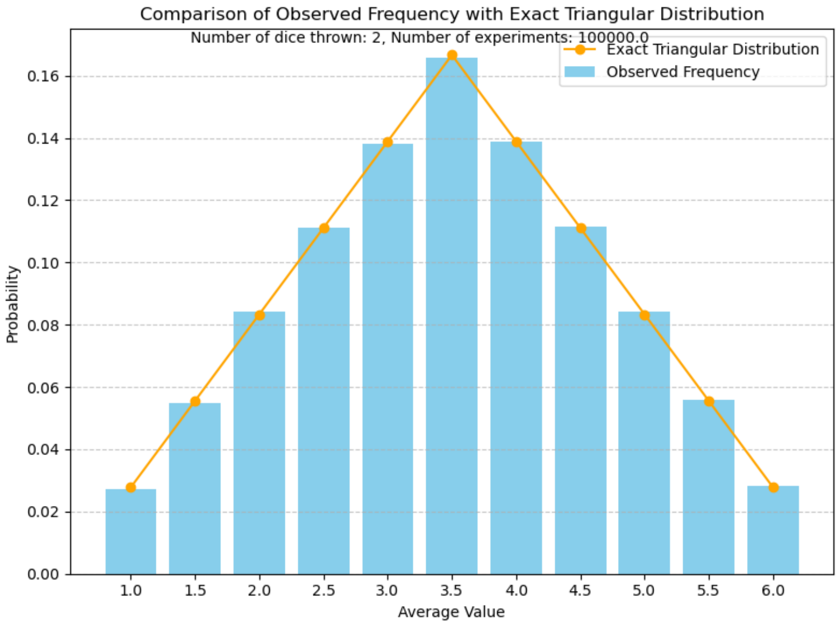

Figure 1.2: Observed frequency versus triangular probability distribution

1.3 Third Task

Consider now the more general case where we throw N_{dice} and calculate the average value

$$x_i = \frac{\sum_{i=1}^{N_{\text{dice}}} \text{rand}(6)}{N_{\text{dice}}}$$

Calculate the probability distribution $p(x)$ and assume that for large N_{dice}, the spacing $dx = 1/N_{\text{dice}}$ between two consecutive values of x is small. Namely, consider that the random variable x is a continuous variable, so its normalization condition is

$$\int p(x)dx = 1$$

Compare this result with the prediction of the **Central Limit theorem**:

$$g(x) = \frac{1}{\sqrt{2\pi}\sigma'} e^{-\frac{(x-\mu)^2}{2\sigma'^2}}$$

where the reduced variance is given by $\sigma' = \sigma/\sqrt{N_{\text{dice}}}$. We will consider that $N_{\text{dice}} = 6$ is a sufficiently large number for the theorem to apply.

1.3.1 Results

- $\mu = 3.5$ vs estimated mean: 3.49945 $\rightarrow$ relative error of 0.00157%.
- $\sigma = 0.69722$ vs estimated std: 0.69601 $\rightarrow$ relative error of 0.1735%.

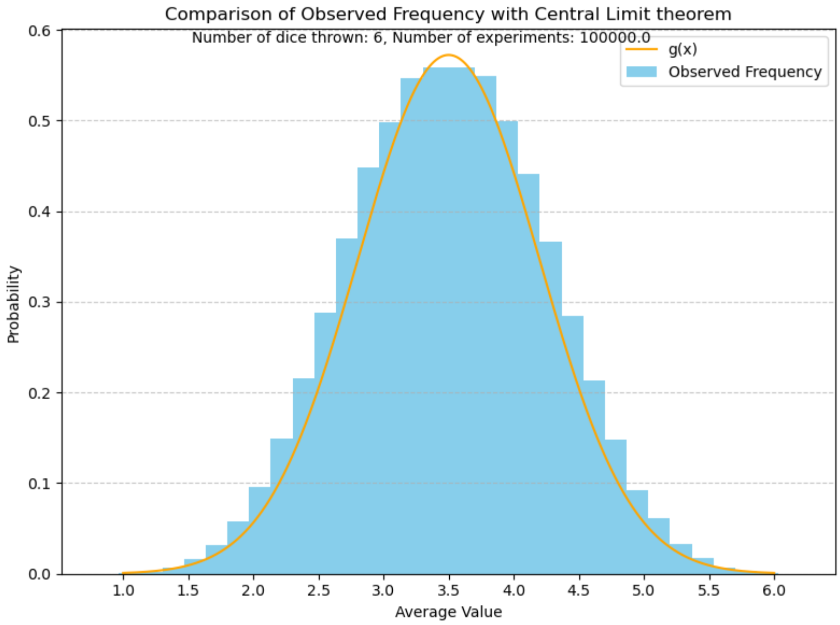

Figure 1.3: Observed frequency versus Central Limit theorem

1.4 Fourth Task

Finally, we are asked to calculate the average value and to estimate the statistical error, associated with such an estimation. Let us assume that we throw a single die to estimate the mean value

$$\mu = \langle x \rangle \approx \frac{\sum_{i=1}^{N_{\text{iter}}} x_i}{N_{\text{iter}}}$$

and the variance

$$\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2 \approx \frac{\sum_{i=1}^{N_{\text{iter}}} x_i^2}{N_{\text{iter}}} - \left(\frac{\sum_{i=1}^{N_{\text{iter}}} x_i}{N_{\text{iter}}} \right)^2$$

Then, calculate this mean value by throwing the die $N_{\text{iter}} = 10$ and $N_{\text{iter}} = 100$ times and estimate the statistical error by $\varepsilon = \sigma / \sqrt{N_{\text{iter}}}$ where the variance is an estimation of the statistical error, i.e., the difference between $\mu(N_{\text{iter}})$ and the exact value of μ.

- For 10 throws of a single die:
  - $\mu = 3.5$ versus estimated mean: 3.9 $\rightarrow$ relative error of 11.429%.
  - $\sigma = 1.7078$ versus estimated std: 1.5780 $\rightarrow$ relative error of 7.600%.
  - Statistical error of: 12.6491%.
- For 100 throws of a single die:
  - $\mu = 3.5$ versus estimated mean: 3.38 $\rightarrow$ relative error of 3.429%.
  - $\sigma = 1.7078$ versus estimated std: 1.8803 $\rightarrow$ relative error of 10.101%.
  - Statistical error of: 1.200%.