STATISTICS 101 - EXERCISE 01

In the lecture, we established that for a random variable X, we can write $X = a$ and mean $X(\omega) = a$, where ω is the set of all possible outcomes of an experiment. Further, the *probability mass function* is given by $p(a) = P(X = a)$, where

$$P(X) = \sum_{\omega \in X} P(\omega).$$

We defined the *cumulative* distribution function as

$$F(a) = P(X \leq a)$$

and the *expected value* is given by

$$E(X) \equiv \mu = \sum_{i=1}^{N} a_i p(a_i).$$

1. Throwing 6-sided dice

Imagine throwing 2 dice n times. We can denote the result of each trial by (i, j), where i and j is the number of respective die that shows upwards.

(1) What is the probability of $M(i, j) = \max(i, j)$, i.e. $p(\max(i, j))$?
  (a) Simulate the process 100 times and plot the PDF/CDF in a *histogram.*
    (i) How many bins should the histogram have?
    (ii) Why is it that theoretical values and simulated ones do not coincide?
    (iii) What happens if you increase the sample size to 10^6?
  (b) What is the expected outcome of the experiment?
  (c) Would anything change if we took the number on the die that faces the table?
(2) Consider the following game:

$$G(i, j) = \begin{cases} €500 & i + j = 7 \\ -€100 & \text{else} \end{cases}$$

  (a) What is the probability of winning a game?
  (b) What is the expected amount of money you win/ loose per game?
  (c) Would you rather play the game 10/100/1000 times? (Simulate it!)
  (d) The standard deviation of a (discrete) random variable is defined as

$$\sigma(X) = \sqrt{\mathrm{var}(X)} = \sqrt{E(X - \mu)^2} = \sqrt{E(X^2) - E(X)^2} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} p(a_i)(a_i - \mu)^2}.$$

  How does this quantity change by playing 10/100/1000 times?
  Would you still play the game?

(e) How can we model this game in terms of a *Binomial* distribution $B(n, p)$? Find analytic expressions for mean and standard deviation.

Note:
$$\sigma(X+Y) = \sqrt{\mathrm{var}(X) + \mathrm{var}(Y) + 2\mathrm{cov}(X,Y)},$$
where $\mathrm{cov}(X,Y) = E(XY) - E(X)E(Y)$.

(f) Plot the analytic and simulated results of mean and standard deviation in dependence of number of games played.

Note: You might need to simulate each case multiple times in order to get a good statistic.

(g) What would be your approach on quantifying the uncertainty of the simulated quantities with respect to sample size?

(3) If given the choice, would you rather play the game in (2) or the following one?
$$K(i,j) = (ij - 10)€$$