Sessions 1 and 2

Exercise 1

Let Y be a random variable that can either be equal to 0 or 1. Let $\mathbb{P}(Y = 1) = p$. Show that $\mathbb{E}(|Y - p|) = 2p(1 - p)$.

Exercise 2: expectation and variance of a random variable following a uniform distribution

Let Y be a random variable taking values in $\{1, 2, 3, ..., N\}$ and such that for any number j belonging to $\{1, 2, 3, ..., N\}$, $\mathbb{P}(Y = j) = \frac{1}{N}$: each possible value of Y is equally likely to get realized.
1) Show that $\mathbb{E}(Y) = \frac{N+1}{2}$. You need to use the fact that $\sum_{j=1}^{N} j = \frac{N(N+1)}{2}$, no need to prove that.

2) Show that $V(Y) = \frac{N^2-1}{12}$. You need to use the fact that $\sum_{j=1}^{N} j^2 = \frac{N(N+1)(2N+1)}{6}$, no need to prove that.

Exercise 3: estimating the variance of the average of n iid binary variables
Let X_1, X_2,...,X_n be n independent and identically distributed random variables. Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ denote the average of the X_is. During the lectures, we have said / will say that we can use $\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$ to estimate the variance of the X_is. Now, assume that for every i in $\{1, ..., n\}$, X_i is either equal to 0 or to 1: the X_is are binary random variables. In such cases, we have said / will say during the lectures that

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \overline{X}(1 - \overline{X}).$$

The goal of this exercise is to prove that formula. Watch out, that formula is true only for binary variables, not for non-binary variables.
1) Show that

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.$$

2) Show that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \overline{X}.$$

3) Use the results of questions 1) and 2) to show that

$$\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \overline{X}(1 - \overline{X}).$$

Exercise 4 (*): proving that if X and Z are independent, then $cov(X, Z) = 0$
Let X be a random variable that can take K values x_1, x_2, ..., x_K. Let Z be a random variable that can take J values z_1, z_2, ..., z_J. Assume that X and Z are independent: for any (x_k, z_j),

$\mathbb{P}(X = x_k, Z = z_j) = \mathbb{P}(X = x_k)\mathbb{P}(Z = z_j)$. In the course, we have seen that if X and Z are independent, then $cov(X, Z) = 0$. The goal of this exercise is to prove that result.

1) Show that $\mathbb{E}(XZ) = \sum_{k=1}^{K} \sum_{j=1}^{J} x_k z_j \mathbb{P}(X = x_k, Z = z_j)$. Hint: what are the values that the random variable XZ can take?

2) Use the result from the previous question, the fact that X and Z are independent, and P1DoubleSum in the slides to show that $\mathbb{E}(XZ) = \mathbb{E}(X)\mathbb{E}(Z)$.

3) Use P1Cov in the slides and the result from the previous question to show that $cov(X, Z) = 0$.