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 $\mathbb{E}(|Y - p|) = 2p(1 - p)$.

Solution The random variable $|Y - p|$ can take two values. It is equal to $|0 - p| = p$ when $Y = 0$, and it is equal to $|1 - p| = 1 - p$ when $Y = 1$. The probability that it is equal to p is equal to $1 - p$ (the probability that $Y = 0$), while the the probability that it is equal to $1 - p$ is equal to p (the probability that $Y = 1$). Therefore, it follows from the definition of the expectation of a random variable that $\mathbb{E}(|Y - p|) = p(1 - p) + (1 - p)p = 2p(1 - p)$.

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.

Solution

$$\begin{aligned}
\mathbb{E}(Y) &= \sum_{j=1}^{N} j\mathbb{P}(Y = j) \\
&= \sum_{j=1}^{N} j\frac{1}{N} \\
&= \frac{1}{N}\sum_{j=1}^{N} j \\
&= \frac{1}{N}\frac{N(N+1)}{2} \\
&= \frac{N+1}{2}.
\end{aligned}$$

The first equality follows from the definition of an expectation. The second follows from the fact $\mathbb{P}(Y = j) = \frac{1}{N}$. The third follows from the second property of the summation operator. The fourth follows from the fact that $\sum_{j=1}^{N} j = \frac{N(N+1)}{2}$.

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.

Solution

We are going to use the fact that $V(Y) = \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2$. That is the first property of the variance operator in the lecture notes. Therefore, to compute $V(Y)$, we first need to compute

$\mathbb{E}(Y^2)$.

$$\begin{aligned}
\mathbb{E}(Y^2) &= \sum_{j=1}^{N} j^2 \mathbb{P}(Y^2 = j^2) \\
&= \sum_{j=1}^{N} j^2 \mathbb{P}(Y = j) \\
&= \sum_{j=1}^{N} j^2 \frac{1}{N} \\
&= \frac{1}{N} \sum_{j=1}^{N} j^2 \\
&= \frac{1}{N} \frac{N(N+1)(2N+1)}{6} \\
&= \frac{(N+1)(2N+1)}{6}.
\end{aligned}$$

The first equality follows from the definition of an expectation. The second follows from the fact that $Y^2 = j^2$ if and only if $Y = j$. Therefore $\mathbb{P}(Y^2 = j^2) = \mathbb{P}(Y = j)$. The third follows from the fact $\mathbb{P}(Y = j) = \frac{1}{N}$. The fourth follows from the second property of the summation operator. The fifth follows from the fact that $\sum_{j=1}^{N} j^2 = \frac{N(N+1)(2N+1)}{6}$.

Then we need to compute $(\mathbb{E}(Y))^2$.

$$\begin{aligned}
(\mathbb{E}(Y))^2 &= \left(\frac{N+1}{2}\right)^2 \\
&= \frac{(N+1)(N+1)}{4}.
\end{aligned}$$

The first equality follows from Question 1.

Finally,

$$\begin{aligned}
V(Y) &= \mathbb{E}(Y^2) - (\mathbb{E}(Y))^2 \\
&= \frac{(N+1)(2N+1)}{6} - \frac{(N+1)(N+1)}{4} \\
&= \frac{(N+1)(2N+1)2}{12} - \frac{(N+1)(N+1)3}{12} \\
&= \frac{(N+1)\left[(2N+1)2 - (N+1)3\right]}{12} \\
&= \frac{(N+1)\left[4N+2-3N-3\right]}{12} \\
&= \frac{(N+1)(N-1)}{12} \\
&= \frac{N^2+N-N-1}{12} \\
&= \frac{N^2-1}{12}.
\end{aligned}$$

The first equality follows from the first property of the variance operator. The second equality follows from the formulae for $\mathbb{E}(Y^2)$ and $(\mathbb{E}(Y))^2$. The other equalities follow from simple algebra.

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

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

Solution

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 &= \frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\overline{X} + \overline{X}^2\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\sum_{i=1}^{n} 2X_i\overline{X} + \frac{1}{n}\sum_{i=1}^{n}\overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\frac{1}{n}\sum_{i=1}^{n} X_i + \frac{1}{n}n\overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}\,\overline{X} + \overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - 2\overline{X}^2 + \overline{X}^2 \\
&= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2.
\end{aligned}$$

The first equality comes from the fact $(a - b)^2 = a^2 - 2ab + b^2$. The second equality follows from the third property of the summation operator. The third equality follows from the first and second properties of the summation operator. The fourth equality follows from the fact $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. The last two equalities follow after some algebra. Notice that this equality holds even when the random variables X_i are not binary. It is the sample analogue of the population formula $V(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$.

2) Show that

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

Solution

As X_i is binary, $X_i^2 = X_i$. Indeed, if $X_i = 0$, $X_i^2 = 0$ as well. And if $X_i = 1$, $X_i^2 = 1$ as well. Therefore,

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

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

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

Solution

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

The first equality follows from question 1), the second follows from question 2).

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?

Solution

XZ can take the values $x_k z_j$ for every k going from 1 to K and for every for every j going from 1 to J. The probability that $XZ = x_k z_j$ is equal to $\mathbb{P}(X = x_k, Z = z_j)$, the probability that $X = x_k$ and $Z = z_j$. Therefore, it follows from the definition of the expectation of a random variable that $\mathbb{E}(XZ) = \sum_{k=1}^{K}\sum_{j=1}^{J} x_k z_j \mathbb{P}(X = x_k, Z = z_j)$.

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

Solution

$$
\begin{aligned}
\mathbb{E}(XZ) &= \sum_{k=1}^{K}\sum_{j=1}^{J} x_k z_j \mathbb{P}(X = x_k, Z = z_j) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{J} x_k z_j \mathbb{P}(X = x_k)\mathbb{P}(Z = z_j) \\
&= \sum_{k=1}^{K}\sum_{j=1}^{J} x_k \mathbb{P}(X = x_k) z_j \mathbb{P}(Z = z_j) \\
&= \left(\sum_{k=1}^{K} x_k \mathbb{P}(X = x_k)\right) \times \left(\sum_{j=1}^{J} z_j \mathbb{P}(Z = z_j)\right) \\
&= \mathbb{E}(X)\mathbb{E}(Z).
\end{aligned}
$$

The first equality follows from the result of question 1). The second one follows from the fact that X and Z are independent. The fourth one follows from P1DoubleSum in the slides. The fifth one follows from the definition of the expectation of a random variable.

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

Solution

$$
\begin{aligned}
cov(X, Z) &= \mathbb{E}(XZ) - \mathbb{E}(X)\mathbb{E}(Z) \\
&= 0.
\end{aligned}
$$

The first equality follows from P1Cov in the slides. The second equality follows from the result of question 2).