1 Vectors and Matrices

I have chosen not to answer this optional question.

2 Derivatives

I have chosen not to answer this optional question.

3 Probability Theory: Sample Space

I have chosen not to answer this optional question.

4 Probability Theory: Properties of Expectation

Let X and Y be two discrete random variables taking values in $\mathcal{X}$ and $\mathcal{Y}$, respectively.

Question 1: Let me now proof that

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y] \tag{1}$$

Let $p_{X,Y}$ be the joint distribution of X and Y, and let p_X and p_Y be the marginal distributions of X and Y, respectively. By definition of the marginalization of a discrete random variable, we know for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ that

$$\sum_{y \in \mathcal{Y}} p_{X,Y}(x,y) = p_X(x) \tag{2}$$

and

$$\sum_{x \in \mathcal{X}} p_{X,Y}(x,y) = p_Y(y) \tag{3}$$

From these statements and the definition of the expectation of a discrete random variable, it follows that

$$\mathbb{E}[X+Y] = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)(x+y) = \tag{4}$$

$$\left(\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)x\right) + \left(\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)y\right) = \tag{5}$$

$$\left(\sum_{x\in\mathcal{X}} x \sum_{y\in\mathcal{Y}} p_{X,Y}(x,y)\right) + \left(\sum_{y\in\mathcal{Y}} y \sum_{x\in\mathcal{X}} p_{X,Y}(x,y)\right) = \tag{6}$$

$$\left(\sum_{x\in\mathcal{X}} x p_X(x)\right) + \left(\sum_{y\in\mathcal{Y}} y p_Y(y)\right) = \mathbb{E}[X] + \mathbb{E}[Y] \tag{7}$$

which proofs the statement on line (1).

Question 2: Assume now that X and Y are independent. Let me now proof that

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \tag{8}$$

By definition of indepence of random variables, we now know for all $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ that

$$p_{X,Y}(x,y) = p_X(x)p_Y(y) \tag{9}$$

From these statements and the definition of the expectation of a discrete random variable, it follows that

$$\mathbb{E}[XY] = \tag{10}$$

$$\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_{X,Y}(x,y)(xy) = \tag{11}$$

$$\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} p_X(x)p_Y(y)xy = \tag{12}$$

$$\sum_{x\in\mathcal{X}} p_X(x)x \sum_{y\in\mathcal{Y}} p_Y(y)y = \tag{13}$$

$$\mathbb{E}[X]\mathbb{E}[Y] \tag{14}$$

which proofs the statement on line (8).

Question 3: Now throw away the assumption that X and Y are independent, and assume that $\mathcal{X} = \{x_1, x_2\}$ and $\mathcal{Y} = \{y_1, y_2\}$. Then this table would provide one option for a valid definition of the joint probability distribution $p_{X,Y}$:

	x_1	x_2	
y_1	0.001	0.199	0.2
y_2	0.499	0.301	0.8
	0.5	0.5	

We see that with this definition of $p_{X,Y}$, then X and Y are not independent, since we for instance have that

$$p_{X_Y}(x_1, y_1) = 0.001 \neq 0.1 = p_X(x_1)p_Y(y_1) \tag{15}$$

We therefore might have - depending on the values of x_1, x_2, y_1 and y_2 - that line (8) does not hold for X and Y. If we for instance set $x_1 = 1000, x_2 = -1, y_1 = 1000$ and $y_2 = -1$, then we have that

$$\mathbb{E}[XY] = 0.001 \cdot 1000000 - 0.199 \cdot 1000 - 0.499 \cdot 1000 + 0.301 \cdot 1 = 302.301 \tag{16}$$

and

$$\mathbb{E}[X]\mathbb{E}[Y] = 0.1 \cdot 1000000 - 0.1 \cdot 1000 - 0.4 \cdot 1000 + 0.4 \cdot 1 = 99500.4 \tag{17}$$

In this case, we clearly see that $\mathbb{E}[XY] \neq \mathbb{E}[X]\mathbb{E}[Y]$.

Question 4: Now go back to the definition of X as an arbitrary discrete, random variable. Let me now proof that

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[X] \tag{18}$$

Let $\mathbb{E}[X]$ be denoted a. Note that a is just some real number. Since p_X is a probability distribution, we know that

$$\sum_{x \in \mathcal{X}} p_X(x) = 1 \tag{19}$$

We now have that

$$\mathbb{E}[\mathbb{E}[X]] = \mathbb{E}[a] = \sum_{x \in \mathcal{X}} p_X(x)a = a \sum_{x \in \mathcal{X}} p_X(x) = a = \mathbb{E}[X] \tag{20}$$

which proofs the statement on line (18).

Question 5: Let me now proof that

$$\mathbb{V}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{21}$$

This we can proof by simply completing the square, and then using our results from question 1 and 3:

$$\mathbb{E}[(X - \mathbb{E}[X])^2] = \tag{22}$$
$$\mathbb{E}[(X^2 + (\mathbb{E}[X])^2 - 2X\mathbb{E}[X]] = \tag{23}$$
$$\mathbb{E}[X^2] + \mathbb{E}[(\mathbb{E}[X])^2] - \mathbb{E}[2X\mathbb{E}[X]] = \tag{24}$$
$$\mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[\mathbb{E}[X]] = \tag{25}$$
$$\mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2\mathbb{E}[X]\mathbb{E}[X] = \tag{26}$$
$$\mathbb{E}[X^2] + (\mathbb{E}[X])^2 - 2(\mathbb{E}[X])^2 = \tag{27}$$
$$\mathbb{E}[X^2] - (\mathbb{E}[X])^2 \tag{28}$$

5 Probability Theory: Complements of Events

Question 1: Let $\mathbb{P}$ be any probability distribution over some sample space Ω, and let $A \subset \Omega$ be some event. By definition of the complement event $\overline{A}$, we know that

$$A \cup \overline{A} = \Omega \tag{29}$$

From the axioms of probability, we also know that

$$\mathbb{P}(\Omega) = 1 \tag{30}$$

and that

$$\mathbb{P}(A \cup \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A}) \tag{31}$$

since A and $\overline{A}$ are mutually exclusive. From these three statements it follows that

$$1 = \mathbb{P}(\Omega) = \mathbb{P}(A \cup \overline{A}) = \mathbb{P}(A) + \mathbb{P}(\overline{A}) \tag{32}$$

from which it follows that

$$\mathbb{P}(A) = 1 - \mathbb{P}(A) \tag{33}$$

which is what I was asked to proof.

Question 2: We flip a fair coin 10 times. What is probability that we observe at least one tail? The complement event of observing at least one tail is only observing heads on all 10 flips. The probability of observing a head on single flip is one half, and since all the flips are independent the probability of observing 10 heads is

$$\left(\frac{1}{2}\right)^{10} \approx 0.000977 \tag{34}$$

The probability of oberserving at least one tail on 10 flips is therefore

$$1 - \left(\frac{1}{2}\right)^{10} \approx 1 - 0.000977 = 0.999023 \tag{35}$$

What is the probability that we observe at least two tails. The complement event is that observe zero or only one tail. From the cumulative binomial distribution we get that the probability of this complement event is approximately 0.010742. Therefore, the probability of observing at least two tails is approximately

$$1 - 0.010742 = 0.989258 \tag{36}$$

6 Digit Classification with Nearest Neighbours

This is some section!