Statistical Foundations of Data Science
Homework 1
Due Monday March 25th

The content of this assignment is based on the lectures on "How to lie with statistics" and "Combinatorics"

1. Determine which of the following are distributions. For those which are not distributions explain why (you shouldn't need more than a sentence for the explanation)

   (a)

   $$p(X) = \begin{cases} \frac{1}{3} & x \in \{1, 6, 20\} \\ 0 & elsewhere \end{cases}$$

   [2]

   Is a distribution

   (b)

   $$p(X) = \begin{cases} \frac{x^2}{10} & x \in \{-2, -1, 1, 2\} \\ 0 & elsewhere \end{cases}$$

   [2]

   Is a distribution

   (c)

   $$p(X) = \begin{cases} \frac{3}{5} & x = 2 \\ \\ \frac{3}{5} & x = 3 \\ \\ \frac{-1}{5} & x = 4 \\ \\ 0 & elsewhere \end{cases}$$

   [2]

   Is NOT a distribution. While the probabilities sum to one negative probabilioties are not allowed.

   (d)

   $$p(X) = \begin{cases} \frac{3}{4^x} & x \in \{1, 2, 3, ...\} \\ 0 & elsewhere \end{cases}$$

   [2]

   Is a distribution

   (e)

   $$p(X) = \begin{cases} \frac{3}{4^x} & x \in \{0, 1, 2, 3, ...\} \\ 0 & elsewhere \end{cases}$$

[2]

Is not a distribution. Probabilities sum to more than 1.

2. One of the two functions given below is a probability density function (i.e. a distribution).

A

$$f(x) = \begin{cases} 3e^{-3x} & x > 0 \\ 0 & elsewhere \end{cases}$$

B

$$f(x) = \begin{cases} 3e^{-4x} & x > 0 \\ 0 & elsewhere \end{cases}$$

(a) Determine which one is a distribution and state why the other one isn't. [3]
A is a distribution. B is not. We compute both integrals.

$$\begin{aligned} \int_0^\infty 3e^{-3x}dx &= 3\frac{1}{3}e^{-3x}|_0^\infty \\ &= e^{-3x}|_0^\infty \\ &= 1 \end{aligned}$$

$$\begin{aligned} \int_0^\infty 3e^{-4x}dx &= 3\frac{1}{4}e^{-4x}|_0^\infty \\ &= \frac{3}{4}e^{-4x}|_0^\infty \\ &= \frac{3}{4} \end{aligned}$$

(b) Compute the mean of the distribution[3]
We compute the mean with by integration by parts or as a gamma integral to be $\frac{1}{3}$

(c) Compute the variance of the distribution[4]
We compute the $\mathbb{E}[X^2]$ again either by integration by parts to be 18, This means that $V(X) = \mathbb{E}[X^2] = \mathbb{E}[X]^2 = \frac{2}{9} - \frac{1}{3^2} = \frac{1}{9}$

3. Let X be distributed continuous uniform between 0 and 10. Compute

a $\mathbb{P}(X > 7)$ [2]
$\frac{3}{10}$ X would need to be 8, 9 or 10.

b $\mathbb{P}(X^2 > 49)$ [3]
$\frac{3}{10}$ This is equivalent to the previous question

c $\mathbb{E}[X]$ [2]
$\mathbb{E}[X] = 5$. This can be seen either by integration or by appeal to the formula for the mean of a continuous uniform.

d $V(X)$ [3]
$\mathbb{E}[X^2] = \frac{100}{3} = 33.3333$. Sp $V(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = 8.3333$

4. Consider the following distribution

$$p(X) = \begin{cases} \frac{1}{5} & x \in \{1,2,3\} \\ c & x = 4 \\ 0 & elsewhere \end{cases}$$

a Find c [3]
$\frac{2}{5}$. It's given by one minus the other probabilities.

b Compute $\mathbb{E}[X]$ [3]
$\mathbb{E}[X] = \frac{1}{5} \times 1 + \frac{1}{5} \times 2 + \frac{1}{5} \times 3 + \frac{2}{5} \times 4 = \frac{14}{5}$.

c Compute $V(X)$ [4]
$\mathbb{E}[X^2] = \frac{1}{5} \times 1 + \frac{1}{5} \times 4 + \frac{1}{5} \times 9 + \frac{2}{5} \times 16 = \frac{46}{5}$. So $V(X) = \frac{46}{5} - (\frac{14}{5})^2 = \frac{34}{25}$

5. A fair coins is tossed 12 times and the number of heads is recorded as X. Compute:

a Compute $\mathbb{E}[X]$ [2]
As the distribution is binomial $\mathbb{E}[X] = np = 12 \times 0.5 = 6$

b Compute $V(X)$ [4]
As the distribution is binomial $V(X) = np(1 = p) = 12 \times 0.5 \times 0.5 = 3$

c $\mathbb{P}(X >= 10)$ [4]

$$\begin{aligned} \mathbb{P}(X >= 10) &= \mathbb{P}(X = 10) + \mathbb{P}(X = 11) + \mathbb{P}(X = 12) \\ &= \frac{\binom{12}{10}}{2^{12}} + \frac{\binom{12}{11}}{2^{12}} + \frac{\binom{12}{12}}{2^{12}} \\ &= \frac{66 + 12 + 1}{4096} \\ &= 0.01928710937 \end{aligned}$$