MTH794P Probability and Statistics for Data Analytics — Week 02 Project

Kavit Tolia

October 9, 2025

Problem 1.

A continuous random variable X has pdf of the form:

$$f_X(x) = \begin{cases} 0 & \text{if } x < \frac{1}{2}, \\ \frac{C}{x^2} & \text{if } x \geq \frac{1}{2} \end{cases}$$

for some constant C.

(a) What is the value of C?

(b) What is $f_X(\frac{1}{2})$?

(c) Find the cdf of X.

(d) Write down a couple of ways you could check your answer to (c) is plausible. Perform those checks and revisit your answer to (c) if necessary.

(e) How would you calculate $\mathbb{P}(1 < X < 3)$ using the cdf?

(f) How would you calculate $\mathbb{P}(1 < X < 3)$ using the pdf?

Solution.

(a) For $f_X(x)$ to be a pdf, we need:

$$\int_{-\infty}^{\infty} f_X(x)\, dx = 1 \Rightarrow \int_{\frac{1}{2}}^{\infty} \frac{C}{x^2}\, dx = 1 \Rightarrow \left[-\frac{C}{x} \right]_{x=\frac{1}{2}}^{\infty} = 1 \Rightarrow C = \frac{1}{2}$$

(b) No, outcomes with the different number of tosses are not equally likely. For example, HH occurs with probability 0.25 whereas TTT occurs with probability 0.125.

(c) The event E "You toss the coin exactly four times" can be written as:

$$E = \{\, TTHH,\ TTHT,\ THTH,\ THTT,\ HTTH,\ HTTT \,\}.$$

(d) We can define the random variable X as "the number of tosses required until the game stops":

$$X : \Omega \to \{2, 3, 4\}, \qquad X(\omega) = \text{number of coin tosses until the experiment stops.}$$

Problem 2. Order of 3 runners in a race

Three runners, Amy, Bea and Cate, take part in a race. The order in which they finish is recorded.

(a) Write down the sample space for this experiment.

(b) Write down the event Amy finishes ahead of Cate as a set.

(c) Write down another event both in words and as a set.

(d) Suppose you had n runners rather than 3. What is the sample space now and how many elements does it contain?

Solution. For the answers, we will denote Amy as A, Bea as B and Cate as C.

(a) The sample space Ω for this space is:

$$\Omega = \{\, ABC,\ ACB,\ BAC,\ BCA,\ CAB,\ CBA \,\}.$$

(b) The event E that Amy finishes before Cate can be written as:

$$E = \{\, ABC,\ ACB,\ BAC \,\}.$$

(c) Another event X can be described as "Bea finishes last":

$$X = \{\, ACB,\ CAB \,\}.$$

(d) If there are n runners, the sample space will be made up of all combinations of the n runners, and it will have $n!$ elements.

Problem 3. Probability with three events

Let A, B, and C be events with

$$\begin{aligned} &\mathbb{P}(A) = 0.7, \quad \mathbb{P}(B) = 0.6, \quad \mathbb{P}(C) = 0.5, \quad \mathbb{P}(A \cap B) = 0.4, \\ &\mathbb{P}(A \cap C) = 0.3, \quad \mathbb{P}(B \cap C) = 0.3, \quad \mathbb{P}(A \cap B \cap C) = 0.2. \end{aligned}$$

Calculate the following:

(a) $\mathbb{P}(A \cup B)$

(b) $\mathbb{P}(A \setminus B)$

(c) The probability that neither of A and B occur.

(d) $\mathbb{P}(A \cup B \cup C)$

(e) The probability that exactly two of A, B, and C occur.

Solution.

(a) $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) = 0.7 + 0.6 - 0.4 = 0.9$

(b) $\mathbb{P}(A \setminus B) = \mathbb{P}(A) - \mathbb{P}(A \cap B) = 0.7 - 0.4 = 0.3$

(c) $\mathbb{P}(\text{neither } A \text{ nor } B) = \mathbb{P}\big((A \cup B)^c\big) = 1 - \mathbb{P}(A \cup B) = 1 - 0.9 = 0.1$

(d) $\mathbb{P}(A \cup B \cup C) = \mathbb{P}(A) + \mathbb{P}(B) + \mathbb{P}(C) - \mathbb{P}(A \cap B) - \mathbb{P}(B \cap C) - \mathbb{P}(A \cap C) + \mathbb{P}(A \cap B \cap C) = 0.7 + 0.6 + 0.5 - 0.4 - 0.3 - 0.3 + 0.2 = 1$

(e) $\mathbb{P}(\text{Exactly 2}) = \mathbb{P}(A \cap B) + \mathbb{P}(B \cap C) + \mathbb{P}(A \cap C) - 3\,\mathbb{P}(A \cap B \cap C) = 0.4 + 0.3 + 0.3 - 3(0.2) = 0.4$

Problem 4. Two random variables

You have a choice of picking one of the two processes, each of which produces a random number. You would prefer a higher number as the outcome. If you pick the first process the outcome is random variable X; if you pick the second process the outcome is random variable Y. The random variables X and Y have pmfs as follows:

n	0	1	2	3
$\mathbb{P}(X = n)$	0.1	0.2	0.3	0.4

n	0	1	2	3
$\mathbb{P}(Y = n)$	0.2	0.2	0.1	0.5

(a) Which procedure would you pick?

(b) Give some reasons to justify your answer.

(c) Suppose that these two processes represent two possible medical procedures to treat a condition, with the numerical outcome being the quality of life of a patient after the treatment. How does this extra context change the way you think about the choice between these procedures?

Solution.

(a) I would pick process X.

(b) $\mathbb{E}(X) = \sum_n n\,\mathbb{P}(X = n) = 0(0.1) + 1(0.2) + 2(0.3) + 3(0.4) = 2$
$\mathbb{E}(Y) = \sum_n n\,\mathbb{P}(Y = n) = 0(0.2) + 1(0.2) + 2(0.1) + 3(0.5) = 1.9$
X has a higher expected value than Y, so I chose X.

(c) Given this context, we also need to understand how the variance behaves for each of these processes.
$\mathbb{E}(X) = 2$, $\mathbb{E}(X^2) = 5 \implies \mathbb{V}(X) = 1$
$\mathbb{E}(Y) = 1.9$, $\mathbb{E}(Y^2) = 5.1 \implies \mathbb{V}(Y) = 1.49$
We can see that X has a higher expected value and a lower variance, so I wouldn't change my choice of picking X as the appropriate process.

Problem 5. Winnings per tossing coins

(a) Consider the following game. You pick an amount of money n. Then we toss a fair coin. If it comes up Heads, I give you £n; if it comes up Tails, you give me £n. We repear the game (you can choose a different amount each time) until you decide to stop.
You decide to adopt the following strategy:

- On the first go stake £1.
- If you win stop.
- If you lose then double your stake on the next game.
- Repeat this (doubling your stake after each loss) until you win.

You argue as follows:

- However many turns the game lasts you will win £1. For example, if it takes 3 turns before the coin comes up Heads, your total gain in pounds is $-1-2+4 = 1$. So you are guaranteed to make £1.

- The game shouldn't last too long. If we let T be the number of times the coin is tossed then you remember (or look it up) the T has a Geometric distribution with parameter 1/2 so $\mathbb{E}(T) = 2$.
- Since in round r you stake £2^{r-1}, the expectation of the amount of money you expect to risk in the final round is $\mathbb{E}(2^{T-1})$ which is also small.

What do you think of each stage of this argument? Is this a sensible strategy to use?

(b) Here is another coin game. We toss a coin until the first time it comes up heads. Support this is on toss number N. If N is even, I pay you £2^N; if N is odd, you pay me £2^N. Let W be your total winnings (which could be negative if you end up paying me!). Find the pmf of the random variable W. What can you say about $\mathbb{E}(W)$?

Solution.

(a) Let's go through each of the stages of the argument in detail here:

- The first stage of this argument works under the assumption of an unlimited bankroll (no cap on losses), no table limits and allowing to stop at the first win. Given these assumptions, you are indeed guaranteed to make £1.
- If T is the number of times the coin is tossed, then T follows a geometric distribution with parameter 1/2. The expected value of a geometric distribution with parameter p is $1/p$. So, we have $\mathbb{E}(T) = 2$. While you could infer that the "game shouldn't last long", we do have to be mindful of the tail risk event of a huge 2^k loss with a rare probability 2^{-k}.
- Let's start with the geometric random variable T, which has $p = 1/2$. We have: $\mathbb{E}(2^{T-1}) = \sum_t 2^{t-1}\mathbb{P}(T = t) = \sum_t 2^{t-1}(1/2)^t = \sum_t (1/2) = \infty$

So, even though the mean of T is 2, the expectation of the final stake is infinte. This is the quantitative reason why the guaranteed £1 is misleading as it requires an infinite bankroll.

(b) The pmf of W is as follows:

w	-2	-2^3	-2^5	-2^7	$\cdots$	2^2	2^4	2^6	$\cdots$
$\mathbb{P}(W = w)$	$\frac{1}{2}$	$\frac{1}{2^3}$	$\frac{1}{2^5}$	$\frac{1}{2^7}$	$\cdots$	$\frac{1}{2^2}$	$\frac{1}{2^4}$	$\frac{1}{2^6}$	$\cdots$

This means the expectation can be written as:
$\mathbb{E}(W) = \sum_n((-1)^n 2^n)(1/2)^n = \sum_n(-1)^n$
This is also known as the Grandi's series and it does not converge. So $\mathbb{E}(W)$ is undefined.
An undefined expectation seems odd, especially given the game seems fair and often ends quickly. However, the possibility of a huge payout or loss (because the stake doubles forever in theory) makes the idea of an expected outcome meaningless. You cannot define a sensible long-run average profit or loss.