BM20A9301 Statistics – Exercise set 4

To be done by 29.1.–2.2.2024

Text in blue or red is not part of the problem or its solution. It's there as extra information to help you learn.

Exercise 1 (Joint distribution)**.** Let S be the set of all students in class, $|S| = 60$, and $H \subseteq S$ is the set of students who have done their homework already, $|H| = 5$. Let's interview two students at random and let

$$X_1 = \begin{cases} 1, & \text{if the first student} \in H \\ 0, & \text{otherwise} \end{cases}$$

$$X_2 = \begin{cases} 1, & \text{if the second student} \in H \\ 0, & \text{otherwise} \end{cases}$$

(a) Assuming that you pick the students with replacement (=the second student is selected from all the students), calculate the joint distribution of (X_1, X_2).

(b) Assuming that you pick the students without replacement (=the second student is selected from the students except the first student), calculate the joint distribution of (X_1, X_2).

(c) Calculate the marginal distributions in both cases.

(d) Are X_1 and X_2 independent in (a)? In (b)?

Exercise 2 (Expected value & variance definition)**.** Calculate the expected values and variances of the random variables A, B, C and D with the following distributions:

k	1	2
$P(A = k)$	0.9	0.1

k	1	2
$P(B = k)$	0.1	0.9

k	6	7
$P(C = k)$	0.9	0.1

k	1	2	3	4
$P(D = k)$	0.59	0.15	0.21	0.05

Exercise 3 (Expected value usage)**.** In Finland, you can buy a lottery ticket called "Ässä–arpa". It costs 4 euros. Every batch has 3 000 000 printed tickets. The amount (in euros) and total number (N) of prize categories are as follows:

Euros	N
100 000	5
2 000	40
1 000	160
500	1 000
30	16 000
20	80 000
10	180 000
5	240 000
4	250 000

What is the expected **net** win in the case of buying one ticket?

Exercise 4 (Simulating a probability)**.** Using Excel, RStudio, Python or whatever you want, estimate the following probability:

We roll three dice. What is the probablity the the largest number is at least 3 higher than the lowest number. For example if the rolls are $(2,5,4)$ then $5-2 \geq 3$ satisfies this, but in the case $(2,4,4)$ it is not true since $4-2=2<3$.

Recall the frequentist interpretation of probability: that probability of an event, is the number of times the event happens divided by the total number of times the random experiment is repeated when the number of repetitions grows very large.

Extra challenge: can you calculate the probability by hand?

Exercise 5 (Distributions)**.** Which distribution will model the following random variables' distribution? For example, if you think that X has the binomial distribution with parameters $n=5$ and $p=0.3$ write $X \sim bin(5,0.3)$. Be ready to explain why. Then use a formula, Excel, distribution tables or RStudio to answer the question. **Hint:** Recall that the slides from lecture 03 listed functions in R which calculate the PDF or CDF of various distributions.

(a) W is the waiting time for the next train in minutes. You know that the train will come at the earliest in 5 minutes but no later than after 10 minutes. That's all you know. What is the probability the bus will arrive before 8 minutes pass?

(b) You maintain a server. Based on its specifications it can handle a maximum of 100 requests every second. Every second, what is the probability that it will crash? The requests R are coming in at around 70 a second on average ($E(R)=70$) and are independent from each other.

(c) You roll a die 60 times and are interested in how many times you roll a six. What is the probability that you roll a six exactly 10 times? What about 10 times or less?

Exercise 6 (Normal distribution)**.** To become a member of the high IQ score association Mensa, you need to have an IQ higher than 98% of the population. The population's mean IQ is defined to be 100. If you score more than 130 in the IQ test you will be offered membership. Assuming that the population's IQ scores are normally distributed, what is the standard deviation?