ISEP – 2nd year – IF.2301 Data Science Fundamentals

EXAM OF PROBABILITIES
DECEMBER 19th 2023

SOLUTIONS

LAST NAME and first name:	
ISEP Number:	
TD group :	G1 / G2

This exam lasts **2 hours**, **any document**, **cell phones, computers and calculators are not allowed**

Instructions :

- Concerning the grading, particular attention will be paid to the quality of writing. Any numeric result without justification will be ignored.
- The last page contains the table of the standard normal distribution.
- Calculators not being allowed, if a calculation seems long or impossible to calculate without a machine, please, leave the general expression with the numerical values (without calculating the final value).

Theoretical questions

Question 1 Let X be a random variable with expected value μ and standard deviation σ_X. Consider the random variable $Y = \frac{X-\mu}{\sigma_X}$.

4. What is the expected value of Y?
5. What is the variance of Y? You will denote this value σ_Y.
6. What is the covariance $cov(X,Y)$ between X and Y ?

Answer :

- $E(Y) = E\left(\frac{X-\mu}{\sigma_X}\right) = \frac{E(X-\mu)}{\sigma_X} = 0.$
- $\sigma_Y^2 = var(Y) = var\left(\frac{X-\mu}{\sigma_X}\right) = \left(\frac{var(X-\mu)}{\sigma_X^2}\right) = \left(\frac{var(X)}{\sigma_X^2}\right) = 1$
- $cov(X,Y) = E(XY) - E(X)E(Y) = E\left(X\frac{X-\mu}{\sigma_X}\right) - 0 = \frac{E(X^2-\mu X)}{\sigma_X} = \frac{E(X^2)-\mu E(X)}{\sigma_X} = \frac{E(X^2)-\mu^2}{\sigma_X} = \frac{\sigma_X^2}{\sigma_X} = \sigma_X$

Question 2 Consider two events A and B associated to a random experiment. Let P(A) and P(B) be the probabilities of A and B respectively.

3. What condition should P(A) and P(B) verify to say that A and B are independent?
4. What condition should P(A) and P(B) verify to say that A and B are disjoint?

Answer :

3. P(A ∩ B) = P(A)P(B)
4. P(A ∩ B) =0 because A ∩ B = ∅

Exercise 1

Consider an archery game with 18 shooters distributed into 4 groups. In the first group there are 5 shooters with a probability of 0.8 of hitting the target. In the second group there are 7 shooters with a probability of 0.7 of hitting the target. In the third one there are 4 shooters with a probability of 0.6 of hitting the target. And in the last group there are 2 shooters with a probability of 0.5 of hitting the target.

a) What is the probability that a shooter randomly chosen misses the target?

b) The target was missed, which group is the shooter most likely to belong to?

Answers :

We define the following events:
a)
E_i : "the chosen shooter belongs to the group i = 1,2,3,4"
F: "when shooting he misses the target";
Therefore, the event F can be written as follows,

$P(F) = \sum_{i=1}^{4} P(E_i)P(F|E_i)$ (total probability law)

$$P(F) = \left(\frac{5}{18}\right) * 0.2 + \left(\frac{7}{18}\right) * 0.3 + \left(\frac{4}{18}\right) * 0.4 + \left(\frac{2}{18}\right) * 0.5 = \frac{1+2.1+1.6+1}{18} = 5.7/18.$$

b) It is necessary to calculate $P(E_i/F)$ for all i and compare the values. Using the Bayes theorem :

$$P(E_i|\, F) = \frac{P(E_i \cap F)}{P(F)} = \frac{P(F|E_i)P(E_i)}{P(F)}$$

- $P(E_1|\, F) = \frac{\left(\frac{5}{18}\right)*0.2}{5.7/18} = 1/5.7$
- $P(E_2|\, F) = \frac{\left(\frac{7}{18}\right)*0.3}{5.7/18} = 2.1/5.7$
- $P(E_3|\, F) = \frac{\left(\frac{4}{18}\right)*0.4}{5.7/18} = 1.6/5.7$
- $P(E_2|\, F) = \frac{\left(\frac{2}{18}\right)*0.5}{5.7/18} = 1/5.7$

So, most likely the shooter comes from group 2.
For this exercise, since the denominator is the same, we just needed to compare the values $P(F|E_i)P(E_i)$ and choose the largest one.

Exercise 2

You choose one of the following integers: 1,2,3,4. After eliminating all the integers (if any) smaller than the chosen one, you choose one of the remaining ones (for example, if the first number is 2, the second choice is made from among the numbers 2,3,4). Let X and Y be the numbers obtained in the first and second elections respectively.

a) What is the joint probability distribution of X and Y ? You can use a contingency table.
b) Determine the marginal probability distributions of X and Y.
c) Calculate the expected value of X and the expected value of Y.
d) Determine the conditional probability function of Y, given X = 3 and the expected value of Y given X=3.
e) Calculate P[Y - X > 0].
f) Are X and Y independent?

Answers:

g) And
h)

		Y				
		y = 1	y = 2	y = 3	y = 4	**P(X=x)**
X	x = 1	1/16	1/16	1/16	1/16	1/4
	x = 2	0	1/12	1/12	1/12	1/4
	x = 3	0	0	1/8	1/8	1/4
	x = 4	0	0	0	1/4	¼
	P(Y=y)	1/16	7/48	13/48	25/48	1

i) E(X) =10/4=5/2 ; E(Y) = (1*3+2*7+3*13+4*25)/48=39/12=3.25 (no need to give the exact value with the decimals.

j)

	Y			
	y = 1	y = 2	y = 3	y = 4
P(Y=y /X = 3)	0	0	1/2	1/2

E(Y/X=3)=3.5

k) P[Y - X > 0] = P(Y > X) to calculate this probability we remark that given the experiment Y is at least equal to X. That means that we can calculate this probability by subtracting the probability of the probability of the complement event to 1, that is, the values on the main diagonal of the contingency table : P(Y>X) =1 – P(X=Y) = 1- (1/16 + 1/12 +1/8 +1/4)= 1 – ((3+4+6+12)/48)=1-25/48= 23/48 very close to 0.5.

l) For X and Y to be independent the following condition should be verified P(X=x; Y =y)= P(X=x)P(Y=y) for all (x,y). It is necessary to consider a counter example that does not verify the condition. By looking at the contingency table P(X=3,Y=2)=0 diferent from P(X=3)P(Y=2)=1/4*7/48
Then, X and Y are not independent.

Exercise 3

In a country, the height in centimeters of women aged 18 to 65 years old can be modeled by a random variable X following the Gaussian distribution with expected value μ_X=165 [cm] and standard deviation σ_X=6[cm]. The height of men aged 18 to 65 years old, can be modeled by a random variable Y following the Gaussian distribution with expected value μ_Y =175[cm] and standard deviation σ_Y =10[cm].

5. What is the probability that the height of a woman chosen at random from this country is between 1.53[m] and 1.77[m]?
6. Determine the probability that a man chosen at random from this country is taller than 1.70 [m].
7. In general, one considers that a man is tall if his height is more than 180cm. We choose 5 women at random. What is the probability law of the number of women that are taller than a man considered *tall* ? We assume that the heights among the female population are independent. In addition, we consider that the total population of women is very large that it can be treated as infinite.
8. **(bonus)** Let's consider women's height are independent from men's height. Give the formula to calculate the probability that a woman chosen at random is taller that a man chosen at random? First, define the probability density function of the random variable (X-Y).

Answers :

Let Z be a standard gaussian random variable.

The purpose is to calculate $P(153 \leqslant X \leqslant 177) = P(-2 \leq Z \leq 2) = P(Z \leq 2) - P(Z \leq -2) = 2 * P(Z \leq 2) - 1 \approx 2 * 0{,}9772 - 1 = 0.9544$

The purpose is to calculate $P(Y \geqslant 170) = P(Z \geq -0.5) = P(Z \leq 0.5) \approx 0{,}6915$.

Let us calculate first the probability that a woman is taller than a man considered tall, that is, $P(X > 180) = P(Z > 2.5) = 1 - P(Z < 2.5) = 1 - 0.9938 = 0.0062$.

Then, let us denote U the number of women taller than 1.80m among $n = 5$. Then U follows a binomial distribution with parameters n = 5 and p = 0.0062

We need to calculate $P(X-Y>0)$.

Since X is Gaussien as well as Y and besides they are independent, then, the vector (X,Y) is Gaussian and as a result any linear combination of its components is gaussian. Then, (X-Y) is also Gaussian with parameters :

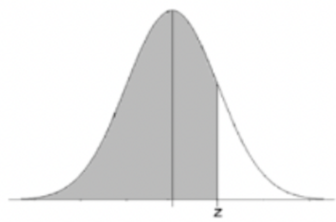
$E(X-Y) = E(X)-E(Y) = -10$cm and variance $V(X-Y) = V(X)+V(Y) = 6^2+10^2 = 136$.

Then, the probability $P(X-Y>0) = P(Z>10/\sqrt{136})$.

CUMULATIVE DISTRIBUTION FUNCTION OF THE STANDARD NORMAL DISTRIBUTION

$$P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du$$

Example $P(Z \leq 1.96) = 0{,}975$



z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998
3,5	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998	0,9998
3,6	0,9998	0,9998	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,7	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,8	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999	0,9999
3,9	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000	1,0000