M902
Βασικές Μαθηματικές Έννοιες
στη
Γλωσσική Τεχνολογία

Project 4

Κυλάφη Χριστίνα-Θεανώ

LT1200012

January, 2021

TABLE OF CONTENTS

Question **1**

i. $n = 3$ independent experiments (coin flips)

$$\Omega = \left\{ \begin{array}{l} \text{ΚΚΚ, ΚΚΓ, ΚΓΚ, ΚΓΓ} \\ \text{ΓΚΚ, ΓΚΓ, ΓΓΚ, ΓΓΓ} \end{array} \right\}$$

ii. $A_1 = \{ \text{ΓΚΚ, ΚΓΚ, ΚΚΓ, ΚΚΚ} \}$
$A_2 = \{ \text{ΓΚΚ, ΚΓΚ, ΚΚΓ} \}$
$A_3 = \{ \text{ΓΚΚ, ΚΓΚ, ΚΚΓ, ΚΚΚ} \} = A_1$
$A_4 = \{ \text{ΚΚΚ, ΓΓΓ} \}$
$A_5 = \{ \text{ΚΚΚ, ΚΚΓ, ΚΓΚ, ΚΓΓ} \}$

Let X a random variable expressing the number of successes (coin flip result → **Κ**), following **Binomial Distribution** (spoilers for *iii* below), $X \sim B(\ 3,\ 0.5)$. Then:

$$P(X = k) = \left(\begin{array}{c} n \\ k \end{array} \right) p^k \ (1-p)^{n-k} \ ,$$

where
p is the **probability** of "success" outcome,
k is the number of **successes**,
n the total number of independent **experiments** performed.

$$P(A_1) = P(X = 2) + P(X = 3) = \left(\begin{array}{c} 3 \\ 2 \end{array} \right) \left(\frac{1}{2} \right)^2 \left(\frac{1}{2} \right) + \left(\begin{array}{c} 3 \\ 3 \end{array} \right) \left(\frac{1}{2} \right)^3 = \frac{3!}{2!1!} \frac{1}{8} + \frac{3!}{3!0!} \frac{1}{8} = = \frac{3}{8} + \frac{1}{8}$$

$$= \frac{N(A_1)}{N(\Omega)} = \frac{4}{8} = 0.5$$

$$P(A_2) = P(X = 2) = \left(\begin{array}{c} 3 \\ 2 \end{array} \right) \left(\frac{1}{2} \right)^2 \left(\frac{1}{2} \right) = \frac{3!}{2!1!} \frac{1}{8} = \frac{3}{8} = \frac{N(A_2)}{N(\Omega)} = 0.375$$

$$P(A_3) = P(A_1) = 0.375$$

$$P(A_4) = P(X = 0) + P(X = 3) = \left(\begin{array}{c} 3 \\ 0 \end{array} \right) \left(\frac{1}{2} \right)^3 + \left(\begin{array}{c} 3 \\ 3 \end{array} \right) \left(\frac{1}{2} \right)^3 = \frac{3!}{0!3!} \frac{1}{8} + \frac{3!}{3!0!} \frac{1}{8} = = \frac{1}{8} + \frac{1}{8}$$

$$= \frac{N(A_1)}{N(\Omega)} = \frac{2}{8} = 0.25$$

Event A_5 concerns only the first coin flip, which is independent of the overall number of experiments. Therefore, the probability of a sole coin flip (the first one) resulting in K, is always $P(K) = \dfrac{1}{2} = 0.5$.

iii. n independent experiments (coin flips)

Here, for event A_2 we apply the same formula as in *ii*, with $X \sim B(\,n,\; 0.5)$:

$$P(A_2) = P(X = 2) = \binom{n}{2}\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^{n-2} = \frac{n!}{2!(n-2)!}\left(\frac{1}{2}\right)^n = \frac{n\,(n-1)}{2}\left(\frac{1}{2}\right)^n$$

As also mentioned in *ii*, the probability of A_5 is always the same and equals the probability of a single coin flip resulting in K, $P(K) = \dfrac{1}{2} = 0.5$.

Question **2**

Let X a random variable following normal distribution $X \sim N(60, 5^2)$ expressing the student weights. Then:

$a)\ P(X > 70) = P\left(\dfrac{X - \mu}{\sigma} > \dfrac{70 - \mu}{\sigma}\right) = P\left(Z > \dfrac{70 - 60}{5}\right) = P(Z > 2) = 1 - P(Z < 2)$

$\qquad = 1 - \Phi(2) = 1 - 0.9772 = 0.0228$

$\beta)\ P(55 < X < 65) = P(X < 65) - P(X < 55) = P\left(\dfrac{X - \mu}{\sigma} < \dfrac{65 - \mu}{\sigma}\right) - P\left(\dfrac{X - \mu}{\sigma} < \dfrac{55 - \mu}{\sigma}\right)$

$\qquad = P\left(\dfrac{X - \mu}{\sigma} < \dfrac{65 - 60}{5}\right) - P\left(\dfrac{X - \mu}{\sigma} < \dfrac{55 - 60}{5}\right) = P(Z < 1) - P(Z < -1)$

$\qquad = \Phi(1) - \Phi(-1) = \Phi(1) - (1 - \Phi(1)) = 2\ \Phi(1) - 1 = 2 * 0.8413 - 1 = 0.6826$

Question **3**

$P(\text{ανοσ}) = 0.7 = p$

The problem can be modelled as a binary outcome (rabbit immunised or not) experiment, executed n times (selecting n rabbits). Then, X is a random variable expressing the number of immunised rabbits picked, with $X \sim B(n, 0.7)$, where $n = 5$:

$i.$ $P(X = 3) = \binom{5}{3} * 0.7^3 * (1 - 0.7)^{5-3} = \dfrac{5!}{3!2!} * 0.7^3 * 0.3^2 = 10 * 0.7^3 * 0.3^2 = 0.3087$

$ii.$ Here, two explanations of the question are going to be followed. However, the resulted probabilities are equal.

1. The probability of picking 3 non-immunised (failure) rabbits and then 1 immunised (success). The task can be modelled as the calculation of the probability that the first success (immunised rabbit) requires k independent trials, thus we calculate the probability of $k - 1$ failures and 1 success (k_{th} trial). In this particular case, X is following the **Geometric Distribution**, $X \sim Geo(0.7)$:

$$P(X = k) = (1 - p)^{k-1}p$$

Then:

$$P(X = 4) = (1 - 0.7)^{4-1}0.7 = 0.3^3 * 0.7 = 0.0189$$

2. The probability of the first rabbit to be the only immunised one, out of 4 rabbits picked in total.

$$P(1_{st} \text{ rabbit immunised }) = 0.7 * (1 - 0.7)^{4-1} = 0.7 * 0.3^3 = 0.0189$$

Question **4**

The solution of this problem, was calculated through code developed in Python. The results are presented below:

		Class	Sentence
Training	1	-	μη χάσετε το χρόνο σας
	2	+	καταπληκτικές ερμηνείες σε ένα δύσκολο έργο
	3	+	η καλύτερη θεατρική παράσταση του χειμώνα
	4	-	δεν ήταν ευχάριστη
	5	+	μια ευχάριστη έκπληξη
Test	1	?	πέρασα μια ευχάριστη θεατρική βραδιά
	2	?	δεν πέρασα μια ευχάριστη θεατρική βραδιά

Word list of concatenated sentences of class " - " :
{ 'καταπληκτικές', 'ερμηνείες', 'σε', 'ένα', 'δύσκολο', 'έργο', 'η', 'καλύτερη', 'θεατρική', 'παράσταση', 'του', 'χειμώνα', 'μια', 'ευχάριστη', 'έκπληξη' }
Count: 15

Word list of concatenated sentences of class " + " :
{ 'μη', 'χάσετε', 'το', 'χρόνο', 'σας', 'δεν', 'ήταν', 'ευχάριστη' }
Count: 8

Word set (union) of the above (all sentences):
{ 'παράσταση', 'σε', 'ερμηνείες', 'καταπληκτικές', 'το', 'δύσκολο', 'ήταν', 'καλύτερη', 'έκπληξη', 'ευχάριστη', 'έργο', 'ένα', 'μη', 'του', 'η', 'μια', 'χρόνο', 'χάσετε', 'δεν', 'θεατρική', 'σας', 'χειμώνα' }
Count: 22

$P(- \,|\, \textbf{test}_1)$

Question **5**

Question **6**

Question **7**

Question **8**

Question **9**

Question **10**