STAT603_Programming_Assignment

October 3, 2023

1 Finding smallest sample size to have a shared birthday with specified probability

[14]:
```
import random
import scipy as sp
from scipy.special import perm
```

2 (a) Consider the Birthday Problem. Modify the Jupyter notebook to calculate using simulations the probability in a group of n people exactly two people share a birthday. Calculate the probability using 1 million simulations for n = 10.

[15]:
```
def Samebirthday(n):
    birthdays=random.choices(list(range(1,366)),k=n) #generates with␣
 ↪replacement n birthdays
    if len(birthdays)!=len(set(birthdays)):
      return True
    else:
      return False
```

[16]:
```
#Simulate a large number of times B=100000 and compute the proportion of
#samples where at least two people share a common birthday

B=1000000
n=10
count=0
for _ in range(B):
    count +=int(Samebirthday(n))


probability=count/B
print(f"Probability that exactly two people share a birthday in a group of {n}␣
 ↪people: {probability:.4f}")
```

```
Probability that exactly two people share a birthday in a group of 10 people:
```

0.1167

3 (b) Derive a formula for the true probability in a group of n people exactly two people share a birthday. Calculate the true probability for n = 10. Compare your answer with the one obtained using simulations in (a)

[17]:
```
def SamebirthdayProbability(n):
    probability_no_shared_birthday = 1.0
    for i in range(n):
        probability_no_shared_birthday *= (365 - i) / 365
    # Calculate the probability that exactly two people share a birthday
    probability_two_shared_birthday = (1 - probability_no_shared_birthday)
    return probability_two_shared_birthday

n = 10
probability = SamebirthdayProbability(n)
print(f"Probability that exactly two people share a birthday in a group of {n}␣
↪people: {probability:.4f}")
```

```
Probability that exactly two people share a birthday in a group of 10 people:
0.1169
```

4 (c) Calculate the minimum number of people in a room for the above probability to exceed 20%.

[21]:
```
def MininumPeople(p):
    n = 2
    while True:
        n += 1
        probability=SamebirthdayProbability(n)
        if probability >=  p:
            #print (n, p, probability)
            return n
```

[22]:
```
# find minimum number of people for the probability to exceed a given␣
↪probability p=0.20
print(f"Minimum number of people for the probability to exceed a given␣
↪probability p=0.20: {MininumPeople(p):.4f}")
```

```
Minimum number of people for the probability to exceed a given probability
p=0.20: 14.0000
```

[23]:
```
%%capture
!pip install nbconvert
!sudo apt-get install texlive-xetex texlive-fonts-recommended␣
↪texlive-plain-generic
```

[ ]:
```
#Mouting the google drive path to read ipynb files
from google.colab import drive
drive.mount('/content/drive')
```

[ ]:
```
# https://saturncloud.io/blog/convert-google-colab-notebook-to-pdf-html/
!jupyter nbconvert  '/content/drive/MyDrive/Colab Notebooks/
↪Ashish_STAT603_Homework3.ipynb' --to pdf
```

[ ]:
```

```

Homework 5 Stat 603

3.43 You take a standard deck of playing cards, and remove one card at random. You then draw a single card. Write S for the event that the card you remove is a six. Write N for the event that the card you remove is not a six. Write R for the event that the card you remove is red. Write B for the event the card you remove is black.

(a) Write A for the event you draw a 6. What is P(A|S)?

Solution: P(A|S) = P (A S)/P(S), P(S) = 1 – 48/52 = 4/52, P(A) = 4/51

P(A|S) = (4/51*4/52)/ (4/52) = 4/51

(b) Write A for the event you draw a 6. What is P (A|N)?

Solution: P(A|N) = P (A N) /P(N), so P(N) = 48/52, P(A) = 4/51

P(A|N) = (4/51 * 48/52)/48/52 = 4/51

(c) Write A for the event you draw a 6. What is P(A)?

Solution: P(A) = 4/51

(d) Write D for the event you draw a red six. Are D and an independent? why?

Solution: For events to be independent, P(A) = P(A|B) = 1/2, P(D) = P(D|A).

P(D) = 2/51, P(A) = 4/51, P(D|A) = P (D A) /P(A) = (2/51 * 4/51)/ (4/51) = 2/51 ==P(D), Hence they are independent events.

(e) Write D for the event you draw a red six. What is P(D)?

Solution: P(D) = 2/51.

3.44 A student takes a multiple-choice test. Each question has N answers. If the student knows the answer to a question, the student gives the right answer, and otherwise guesses uniformly and at random. The student knows the answer to 70% of the questions. Write K for the event a student knows the answer to a question and R for the event the student answers the question correctly.
(a) What is P(K)?

Solution: P(K) = 70/100 = 0.7
(b) What is P(R|K)?

Solution: P(R|K) = P (R K) /P(K) = 1
(c) What is P(K|R), as a function of N?

Solution: Using Bayes' probability function,

P(K/R) = P (K and R)/P (R) = P(K) * P(R|K) / (P(K) * P(R|K) + P(K') P(R|K')

= 0.7*1 / (0.7* 1 + (1-0.7) * 1/N)

= 0.7/ (0.7 + 0.3/N)

(d) What values of N will ensure that P(K|R) > 99%?

Solution: 99/100 < 0.7/ (0.7 + 0.3/N)

= 0.99 < 0.7/(1/N)

= N> 42.86 == 43

3.Pollution of the rivers in the United States has been a problem for many years. Consider
the following events:

A: the river is polluted

B: a sample of water tested detects pollution

C: fishing is permitted

Assume that $P(A) = 0.3$, $P(B|A) = 0.75$, $P(B|A') = 0.20$, $P(C|(A \cap B)) = 0.20$
$P(C|(A^c \cap B)) = 0.15$, $P(C|(A \cap B')) = 0.80$, and $P(C|(A^c \cap B')) = 0.90$.

(a) Find $P(A \cap B \cap C)$.

Solution: $P(A \cap B \cap C) = P(C|A \text{ and } B) * P(A \text{ and } B)$

$= P(C|A \text{ and } B) * P(A)*(B|A)$

$= 0.2 * 0.75*0.3 = 0.045$

(b) Find $P(B' \cap C)$.

Solution: Since $B' \cap C = (A \cap B' \cap C) \cup (A' \cap B' \cap C)$

$P(B' \cap C) = P(A \cap B' \cap C) + P(A' \cap B' \cap C)$

But $P(A \cap B' \cap C) = P(C|A \cap B') * P(A \cap B')$

$= 0.8*(1 - 0.75) * 0.3 = 0.06$

And $P(A' \cap B' \cap C) = 0.9 *(1 - 0.2) *(1 - 0.3) = 0.504$

$P(B' \cap C) = 0.06+0.504 = 0.564$

(c) Find $P(C)$.

Solution: $C = (A \cap B' \cap C) \cup (A' \cap B' \cap C) \cup (A \cap B \cap C) \cup (A' \cap B \cap C)$

$P(C) = (A \cap B' \cap C) + (A' \cap B' \cap C) + (A \cap B \cap C) + (A' \cap B \cap C)$

But $P(A' \cap B \cap C) = P(C|A' \cap B) * P(A' \cap B)$

$= 0.15 * P(B|A') * P(A')$

$= 0.15 * 0.20 * 0.07 = 0.021$

Hence $P(C) = 0.06+0.504+0.045+0.021 = 0.63$

(d) Find the probability that the river is polluted given that fishing is permitted and the sample tested did not detect pollution

Solution: The required probability $= P(A|C \cap B') = P(A \cap B' \cap C)/P(C \cap B') = 0.06/0.546$

$= 0.106$