Faculty of Computer & Information Sciences – Ain Shams University
Statistical Analysis and Apllications

Sheet 4 (Revision on Probability)

1. A researcher hypothesizes that it is possible to detect membrane proteins using the fraction of hydrophobic residues alone. To test this model, the researcher creates a library of 7500 proteins and scores each of these proteins based on their fraction of hydrophobic residues and whether they are membrane proteins. The results of this analysis are shown below:

	Majority Hydrophobic	Majority Hydrophilic
Membrane Bound	2911	961
Cytosolic	713	2915

Given this information, find the likelihood that a novel protein that is primarily hydrophobic is also a membrane protein.

2. A test for a rare disease claims that it will report a positive result for 99.5% of people with the disease, and will report a negative result for 99.9% of those without the disease. We know that the disease is present in the population at 1 in 100,000. Knowing this information, what is the likelihood that an individual who tests positive will actually have the disease?

3. Three jars contain colored balls as described in the table below:

Jar #	Red	White	Blue
1	3	4	1
2	1	2	3
3	4	3	2

One jar is chosen at random and a ball is selected. If the ball is red, what is the probability that it came from the 2nd jar?

4. All tractors made by a company are produced on one of three assembly lines, named Red, White, and Blue. The chances that a tractor will not start when it rolls off of a line are 6%, 11%, and 8% for lines Red, White, and Blue, respectively. 48% of the company`s tractors are made on the Red line and 31% are made on the Blue line. What fraction of the company's tractors do not start when they roll off of an assembly line?

Bonus question: What is the probability that a tractor came from the red company given that it was defective?

5. There are two urns containing colored balls. The first urn contains 50 red balls and 50 blue balls. The second urn contains 30 red balls and 70 blue balls. One of the two urns is randomly chosen (both urns have probability 50% of being chosen) and then a ball is drawn at random from one of the two urns. If a red ball is drawn, what is the probability that it comes from the first urn?

6. An economics consulting firm has created a model to predict recessions. The model predicts a recession with probability 80% when a recession is indeed coming and with probability 10% when no recession is coming. The unconditional probability of falling into a recession is 20%. If the model predicts a recession, what is the probability that a recession will indeed come?

7. Alice has two coins in her pocket, a fair coin (head on one side and tail on the other side) and a two-headed coin. She picks one at random from her pocket, tosses it and obtains head. What is the probability that she flipped the fair coin?

8. In a certain population, 30% of the persons smoke and 8% have a certain type of heart disease. Moreover, 12% of the persons who smoke have the disease.

   a) What percentage of the population smoke and have the disease?
   b) What percentage of the population with the disease also smoke?
   c) Are smoking and the disease positively correlated, negatively correlated, or independent?

9. A company has 200 employees: 120 are women and 80 are men. Of the 120 female employees, 30 are classified as managers, while 20 of the 80 male employees are managers. Suppose that an employee is chosen at random.

   a) Find the probability that the employee is female.
   b) Find the probability that the employee is a manager.
   c) Find the conditional probability that the employee is a manager given that the employee is female.
   d) Find the conditional probability that the employee is female given that the employee is a manager.
   e) Are the events *female* and *manager* positively correlated, negatively correlated, or indpendent?

SUBJECT: DATE / /

: Sheet 4 :

III. II $N = 7500$

	Majority Hydrophobic	Majority Hydrophilic
Membrane Bound	$2911 = P(M \cap H)$	$961 = P(M \cap Hy)$
cytosolic	$713 = P(C \cap H)$	$2915 = P(C \cap Hy)$

Given This information, find The Likelihood That a novel protein That is primarily hydrophobic is also a membrance protein?

Likelihood ≡ Conditional Probability

let
H : hydrophobic , Hy : Hydrophilic
M : membrane Bound , C : Cytosolic

We Want to find The Probability That a novel protein is a primarily hydrophobic (1) is also a membrance (2) protein

$$P(M|H) = \frac{P(M \cap H)}{P(H)}$$

$$P(M \cap H) = \frac{2911}{7500}, \quad P(H) = P(M \cap H) + P(C \cap H)$$

$$= \frac{2911}{7500} + \frac{713}{7500} = \frac{3624}{7500}$$

$$\therefore P(M|H) = \frac{\frac{2911}{7500}}{\frac{3624}{7500}} = \frac{2911}{3624} = 0.8032$$ ✱

OBJECT: .... DATE / /

2.

99.5 % → +ve result with disease.

// عندهم المرض أكثر ناس طلعوا النتيجة +ve

$P(+ve \mid disease) = 0.995$

99.9% → −ve result without disease.

لا يوجد لديهم المرض طلعوا النتيجة −ve

$P(-ve \mid without\ disease) = 0.999$

disease is present in The Population at 1 in 100000

$P(disease\ appearance) = \frac{1}{100\,000} = 0.00001$

Knowing This information, what is The probability That an individual who tests positive will actually have The disease?

we want to find Probability That a patient has The disease given That a +ve Test result.

$$P(with\ disease \mid +ve) = \frac{P(+ve \cap with\ disease)}{P(+ve)} \rightarrow ①$$

$$P(+ve \mid disease) = \frac{P(+ve \cap disease)}{P(disease)}$$

$$P(+ve \cap with\ disease) = P(+ve \mid disease)\, P(disease)$$

$$= (0.995)(0.00001)$$

$$= 0.00000995 \rightarrow ②$$

SUBJECT: DATE / /

$P(+ve) = P(+ve \cap \text{with disease}) + P(+ve \cap \text{without disease})$

$= P(+ve \mid \text{with disease})\,P(\text{disease}) + P(+ve \mid \text{without disease})\,P(\text{without dis})$

$= P(+ve \mid \text{with disease})\,P(\text{disease}) + (1 - P(-ve \mid \text{without disease}))(1 - P(\text{disease}))$

$= (0.995)(0.00001) + (1 - 0.999)(1 - 0.00001)$

$= 0.00000995 + (0.001)(0.99999)$

$= 0.00000995 + 0.00099999$

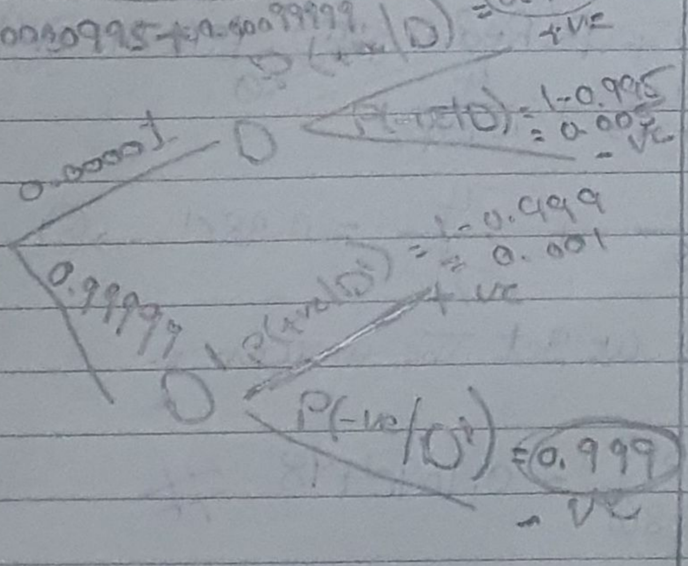
$= 0.00100994 \rightarrow (3)$

From (2) & (3) In (1) we get

$P(\text{with disease} \mid +ve) = \dfrac{0.00000995}{0.00100994} = 0.009852070 4$

$P(D \mid +ve) = \dfrac{P(+ve \cap D)}{P(+ve)} = \dfrac{P(+ve \mid D)\,P(D)}{P(+ve \mid D)\,P(D) + P(+ve \mid D')\,P(D')}$

$= \dfrac{0.00000995}{0.00000995 + 0.00099999} =$



$\cap$

$P(D \cap +ve) = 0.00001 \times 0.995 = 0.00000995$

$P(D \cap -ve) = 0.00001 \times 0.005 = 0.00000005$

$P(D' \cap +ve) = 0.99999 \times 0.001 = 0.00099999$

$P(D' \cap -ve) = 0.9999 \times 0.999 = 0.9989001$

DATE / /

OBJECT:

3.

Jar	Red	White	Blue	Total
1	3	4	1	8
2	1	2	3	6
3	4	3	2	9

one Jar is chosen at random and a ball is selected. if The ball is red, what is The probability That it come from The 2nd Jar?

if ① Then ②

$P(2^{nd}\text{ Jar} \mid Red) = \frac{P(2^{nd}\text{ Jar} \cap Red)}{P(Red)} \rightarrow ①$

$P(2^{nd}\text{ Jar} \cap Red) = \frac{1}{1+2+3} = \frac{1}{6} = 0.1667 \rightarrow ②$

$P(Red) = P(Red \cap 1^{st}\text{ Jar}) + P(Red \cap 2^{nd}\text{ Jar}) + P(Red \cap 3^{rd}\text{ Jar})$

$= \frac{3}{3+4+1} + \frac{1}{1+2+3} + \frac{4}{4+3+2}$

$= \frac{3}{8} + \frac{1}{6} + \frac{4}{9} = 0.9861 \rightarrow ③$

From ③ and ② In ① we get The answer :-

$P(2^{nd}\text{ Jar} \mid Red) = \frac{0.1667}{0.9861} = 0.1690498$ #

SUBJECT:

"4."

Three assembly Lines, (Red, White, Blue)

- 48 % That The Company's tractors are made on The Red Lines.

P(Red) = 0.48

- 31 % The Company's tractors are made on The Blue lines

P(Blue) = 0.31

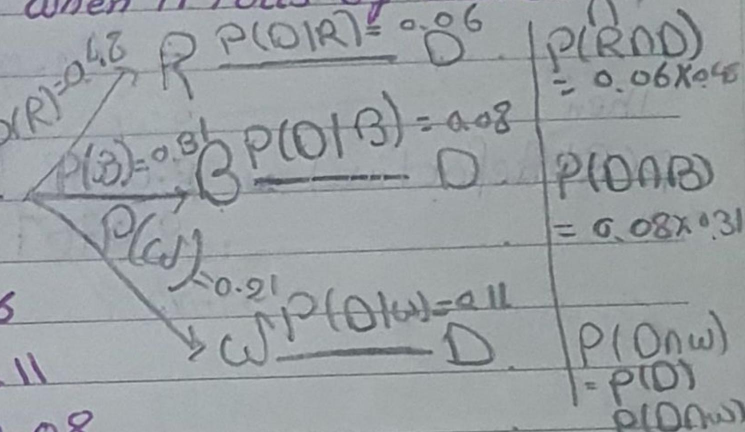
given P(Red) and P(Blue) Then P(White) = 1 − P(Blue) − P(Red) = 1 − 0.48 − 0.31 = 0.21

- a tractor will not start when it rolls of a line.

6 % for lines Red

Defective 11 % for lines White

8 % for lines Blue.



P(not start | Red) = 0.06

P(not start | White) = 0.11

P(not start | Blue) = 0.08

What fraction of The Company's tractors do not start when They roll of an assembly line? P(D)

Here we want to find The Probability That not start at any assembly line

P(not start) = P(not start ∩ Red) + P(not start ∩ White) + P(not start ∩ Blue)

= P(not start | Red) P(Red) + P(not start | White) P(White) + P(not start | Blue) P(Blue).

OBJECT: .... DATE / /

$P(\text{not start}) = (0.06)(0.48) + (0.11)(0.21) + (0.08)(0.31)$
$= 0.0288 + 0.0231 + 0.0248$
$= 0.0767$ ##

What is The probability That a tractor came from The red Company given That it was defective?

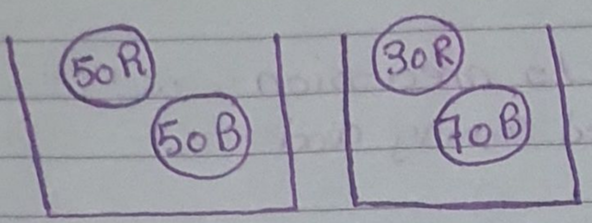
$$P(R \mid \text{defective}) \equiv P(R \mid \text{not start}) = \frac{P(R \cap \text{not start})}{P(\text{not start})}$$

$P(R \cap \text{not start}) = P(\text{not start} \mid R)\, P(R)$
$= (0.06)(0.48) = 0.0288$

Then

$$P(R \mid \text{defective}) = \frac{0.0288}{0.0767} = 0.3755$$ ##

5.



urn (A) urn (B)

- one of Two urns selected
- probability of selected urn A = Probability of selected urn B = 0.5.
- after select one of Two urns a ball is drawn at random from one of Two urns.

if a red ball is drawn, what is The probability That it comes from The first urn?

$$P(\text{urn } A \mid \text{red}) = \frac{P(\text{urn } A \cap \text{red})}{P(\text{red})} \rightarrow (1)$$

$$P(\text{urn } A \cap \text{red}) = \frac{50}{100} = 0.5 \rightarrow (2)$$

$$P(\text{Red}) = P(\text{urn } A \cap \text{red}) + P(\text{urn } B \cap \text{red})$$

$$= \frac{50}{100} + \frac{30}{100} = \frac{80}{100} = 0.8 \rightarrow (3)$$

From (3), (2) in (1) we get :-

$$P(\text{urn } A \mid \text{red}) = \frac{0.5}{0.8} = 0.625$$

OBJECT: DATE / /

when

16.1

- let :- P: Predicts a recession
  I: indeed coming

- 80% when a recession is indeed coming The model predicts a recession

$P(P|I) = 0.8$

- 10% when no recession is coming. The model predicts a recession.

$P(P|I') = 0.1$

- unconditional probability of falling into a recession is 20%

$P(I) = 0.2$ = indeed coming

What is The probability That a recession will indeed come?

$$P(I|P) = \frac{P(I \cap P)}{P(P)} \rightarrow ①$$

$P(I \cap P) = P(P|I)\,P(I) = (0.8)(0.2) = 0.16 \rightarrow ②$

$P(P) = P(I \cap P) + P(I' \cap P)$

$= P(P|I)P(I) + P(P|I')P(I')$

$= (0.8)(0.2) + (0.1)(1-0.2) = 0.16 + 0.08$

$= 0.24 \rightarrow ③$

From ③ & ② in ① we get :- $P(I|P) = \frac{0.16}{0.24} = 0.67$

sema

SUBJECT: .... DATE / /

// 7. //

- Two Coins:
  - → fair coin (head on one side and Tail on The other side)
  - → Two - headed coin
- She picks one at random
- tosses it and obtains head.

What is The probability That she flipped The fair coin. $P(\text{fair}|\text{head})$

∴ tosses The Coin she picked it and obtains head
This may be Come from fair Coin and unfair Coin.

$P(H \mid \text{fair coin}) = 1/2$	$P(\text{select fair coin}) = P(\text{select unfair coin})$
$P(H \mid \text{unfair coin}) = 1$	$= 0.5$

we want to find

$$P(\text{fair coin} \mid \text{head}) = \frac{P(\text{fair coin} \cap \text{head})}{P(\text{head})} \rightarrow ①$$

$$P(\text{fair coin} \cap \text{head}) = P(H \mid \text{fair coin})\, P(\text{fair coin})$$
$$= (0.5)(0.5) = 0.25 \rightarrow ②$$

$$P(H) = P(\text{fair coin} \cap H) + P(\text{unfair coin} \cap H)$$
$$= P(H \mid \text{fair coin})\, P(\text{fair coin}) + P(H \mid \text{unfair coin})\, P(\text{unfair})$$
$$= (0.5)(0.5) + (0.5)(1)$$
$$= 0.25 + 0.5 = 0.75 \rightarrow ③$$

From ③, ② in ① we get :-

$$P(\text{fair coin} \mid \text{head}) = \frac{0.25}{0.75} = 0.33 \quad \#$$

8.

- 30 % of The persons smoke.

$P(\text{smoke}) = 0.3$

- 8 % have a certain type of heart disease

$P(\text{heart disease}) = 0.8$

- 12 % of the persons who smoke have The disease
  who (1) have (2)

$P(\text{disease} \mid \text{smoke}) = 0.12$

(a) What percentage of The population smoke and have The disease ?

$P(\text{Smoke} \cap \text{disease}) = P(\text{disease} \mid \text{smoke})\, P(\text{smoke})$

$= (0.12)(0.3) = 0.036$

$= 3.6\%$ smoke and have The disease.

(b) What percentage of The population with The disease also smoke ?
with (1) also (2)

$P(\text{smoke} \mid \text{disease}) = \dfrac{P(\text{smoke} \cap \text{disease})}{P(\text{disease})} = \dfrac{0.036}{0.8}$

$= 0.045$

$= 4.5\%$

(c) Are smoking and The disease positively correlated, negatively correlated, or independent ?

$P(\text{smoke} \cap \text{disease}) = 0.036$

$P(\text{smoke})\, P(\text{disease}) = (0.3)(0.8) = 0.24$

$P(\text{smoke} \cap \text{disease}) < P(\text{smoke})\, P(\text{disease})$

∴ smoke, disease are negtively correlated //

19.1

- $N = 200$
- number of women $= 120$
  number of men $= 80$
- 120 female employees, 30 are classified as managers, while 20 of the 80 male employees are managers.

200 → 120 F → 30 managers, 90 others
200 → 80 M → 20 managers, 60 others

(a) Find the probability that the employee is female.

$$P(\text{Female}) = \frac{120}{200} = 0.6$$

(b) Find the probability that the employee is a manager.

$$P(\text{manager}) = P(\text{manager} \cap \text{Female}) + P(\text{manager} \cap \text{male})$$
$$= \frac{30}{200} + \frac{20}{200} = 0.15 + 0.1 = 0.25$$

(c) Find the conditional probability that the employee is a manager given that the employee is female.

$$P(\text{manager} \mid \text{Female}) = \frac{P(\text{manager} \cap \text{Female})}{P(\text{Female})} = \frac{30/200}{120/200} = \frac{30}{120}$$
$$= 0.25$$

(d) Find The conditional probability That The employee is female given That The employee is a manager.

$$P(\text{female} \mid \text{manager}) = \frac{P(\text{female} \cap \text{manager})}{P(\text{manager})} = \frac{30/200}{0.25}$$

$$= \frac{0.15}{0.25} = 0.6$$

(e) are The events female and manager positively correlated, negatively correlated, or independent ?

$$P(\text{female} \cap \text{manager}) = \frac{30}{200} = 0.15$$

$$P(\text{female})\,P(\text{manager}) = \frac{120}{200} \cdot 0.25 = (0.6)(0.25)$$

$$= 0.15$$

Then $P(\text{female} \cap \text{manager}) = P(\text{female})\,P(\text{manager})$

female and manager are independents.