W203 Statistics for Data Science

Aditya Mengani

aditya.mengani@berkeley.edu

Aditya Mengani

Monday, 4pm

Summer 2020

Unit 1 Homework

May 3, 2020

1. Gas Station Analytics At a certain gas station, 40% of customers use regular gas (event R), 35% use mid-grade (event M), and 25% use premium (event P). Of the customers that use regular gas, 30% fill their tanks (Event F). Of the customers that use mid-grade gas, 60% fill their tanks, while of those that use premium, 50% fill their tanks. Assume that each customer is drawn independently from the entire pool of customers.

1. (a) What is the probability that the next customer will request regular gas and fill the tank? (3 points)
2. (b) What is the probability that the next customer will fill the tank? (3 points)
3. (c) Given that the next customer fills the tank, what is the conditional probability that they use regular gas? (3 points)

Given :- Let 'N' be the no. events in Ω

$R = \{$ Regular gas is purchased $\}$

$M = \{$ Midgrade gas is purchased $\}$

$P = \{$ Premium gas is purchased $\}$

Then, $P(R) = 0.4 \qquad P(M) = 0.35 \qquad P(P) = 0.25$

once, a type of gas is choosen, the second stage involves whether the customer fills their tanks or not

That is represented as,

$F = \{$ Fill the tank $\}$

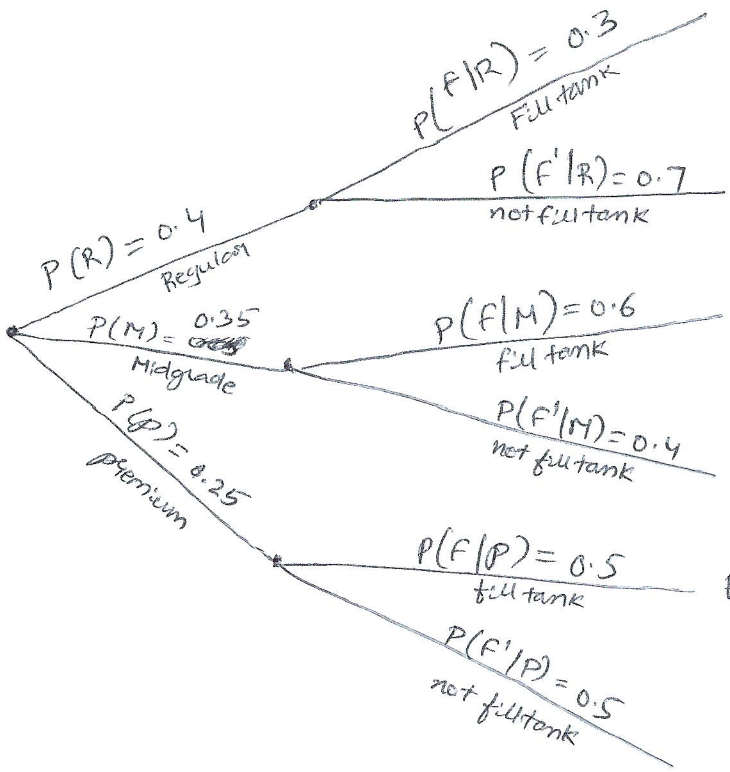
$F' = \{$ doesn't fill the tank $\}$

from the given information in question, it is implied that,

$P(F|R) = 0.3, \quad P(F|M) = 0.6, \quad P(F|P) = 0.5$

I have represented this situation using tree diagram in the next page :

(2)



$P(F|R) \cdot P(R) = P(F \cap R) = 0.12$
$\longrightarrow$ (a1)

$P(F|M) \cdot P(M) = P(F \cap M) = 0.21$
$\hookrightarrow$ (b1)

$P(F|P) \cdot P(P) = P(F \cap P) = 0.25 \times 0.5$
$= 0.125$
$\llcorner$ (c1)

(a). The probability that next customer will request regular gas and fill the tank is represented as:

$P(R \cap F)$

According to multiplication Rule,
$P(R \cap F)$ can be represented as,
$P(R \cap F) = P(F|R) \cdot P(R)$

According to tree diagram above, the value 0.12 can be substituted from (a1),

hence, $P(R \cap F) = 0.4 \times 0.3 = 0.12$

Thus, the probability,

$P\left(\begin{array}{l}\text{customer request regular gas and} \\ \text{fills tank}\end{array}\right) = \underline{\underline{0.12}}$ —(d1)

(b)

(b). The solution to this problem, can be derived from "Total law of probability"

which states, Let A1, A2 .. Ak be M.E. events, then for event B which is another event

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(B_2) + \dots P(B|A_K)(A_K)$$

$$= \sum_{i=1}^{K} P(B|A_i) \cdot P(A_i)$$

where $A_1, A_2 \dots A_K$ are mutually enclusive events

In our problem R, M, P are mutually exclusive events

thus for an other Event 'F' which is whether customer will fill the tank can be expressed as,

$$P(F) = \underset{\substack{\downarrow \\ \text{(derived from tree diagram a1)}}}{P(F|R) \cdot P(R)} + \underset{\substack{\downarrow \\ \text{(derived from tree diagram b1)}}}{P(F|M) \cdot P(M)} + \underset{\substack{\downarrow \\ \text{(derived from tree diagram c1)}}}{P(F|P) \cdot P(P)}$$

The values have been derived earlier in tree diagram, which can be represented as

$$P(F) = (a1) + (b1) + (c1)$$

$$= 0.12 + 0.21 + 0.125$$

$$P(F) = \underline{0.455} \quad \text{———— (e1)}$$

Thus, $P\left(\text{probability that the next customer will fill a tank}\right) = 0.455$

(c). The solution to this problem can be derived using Baye's theorem. (4)

It states, let $A_1, A_2 \ldots A_n$ be a collection of events with prior probability $P(A_i)$ and are mutually exclusive and exhaustive, then for any given other event B for which $P(B) > 0$, the posterior probability of A_j, given that B occured is

$$\boxed{P(A_j|B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B|A_j) \cdot P(A_j)}{\sum_{i=1}^{K} P(B|A_i) \cdot P(A_i)} \quad \text{where } j = 1 \ldots K}$$

In our problem statement $A_j = \{R, M, P\}$

$B = F = \{\text{Event that a tank is filled}\}$

So, probability P (Given next customer fills tank, probability that they use regular gas)

can be written as $P\left(\frac{R}{F}\right)$

using formula of Bayes' theorem,

$$P\left(\frac{R}{F}\right) = \frac{P(R \cap F)}{P(F)} = \frac{P(R|F) \cdot P(F)}{P(F)}$$

from earlier calculations, (d1) and (e1),

we know $P(R \cap F) = 0.12$ and $P(F) = 0.455$

Substituting them $P(R|F) = \frac{0.12}{0.455} = 0.264$

Thus, P (Given next customer fills a tank, probability that they use regular gas) $= \underline{\underline{0.264}}$

2. The Toy Bin

In a collection of toys, 1/2 are red, 1/2 are waterproof, and 1/3 are cool. 1/4 are red and waterproof. 1/6 are red and cool. 1/6 are waterproof and cool. 1/6 are neither red, waterproof, nor cool. Each toy has an equal chance of being selected.

1. (a) Draw an area diagram to represent these events. (3 points)
2. (b) What is the probability of getting a red, waterproof, cool toy? (3 points)
3. (c) You pull out a toy at random and you observe only the color, noting that it is red. Conditional on just this information, what is the probability that the toy is not cool? (3 points)
4. (d) Given that a randomly selected toy is red or waterproof, what is the probability that it is cool? (3 points)

(a). Let 'Ω' be the space of all events then,

R = {toys that are red}

W = {toys that are waterproof}

C = {toys that are cool}

$R \cap W$ = {toys that are red and waterproof}

$R \cap C$ = {toys that are red and cool}

$W \cap C$ = {toys that are waterproof and cool}

Z = {set of toys that are neither red, waterproof, nor cool}

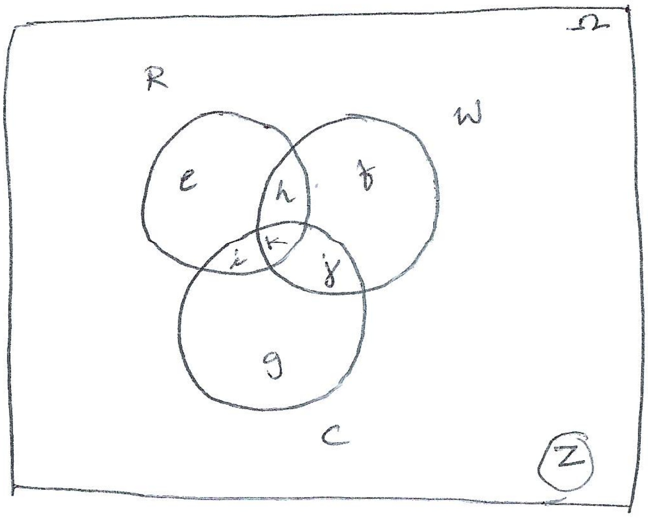
Then, it's given,

$P(R) = 1/2$ $P(W) = 1/2$ $P(C) = 1/3$ $P(R \cap W) = 1/4$ $P(R \cap C) = 1/6$

$P(W \cap C) = 1/6$ $P(Z) = 1/6$

$P(\Omega) = 1$ (the entire event space)

This can be represented in a diagram as follows in next page:



From the area diagram above,

$P(R) = \text{Area}(e+h+i+k) = 1/2$; $P(W) = \text{Area}(h+f+k+j) = 1/2$

$P(C) = \text{Area}(i+j+k+g) = 1/3$; $P(R \cap W) = \text{Area}(h+k) = 1/4$

$P(R \cap C) = \text{Area}(i+k) = 1/6$; $P(W \cap C) = \text{Area}(k+j) = 1/6$

$P(Z) =$ Area not in R, W, C $= P(!(R \cup W \cup C)) = 1/6$

Problem (b):- Find $P(R \cap W \cap C)$: It is nothing but area under $\{k\}$ above

from axioms of probability:

$$\boxed{P(R \cup W \cup C) = P(R) + P(W) + P(C) - P(R \cap W) - P(W \cap C) - P(C \cap R) + P(R \cap W \cap C)}$$

$$\Rightarrow P(R \cap W \cap C) = \cancel{P(R)+P(W)+P(C)} \; P(R \cup W \cup C) - [P(R) + P(W) + P(C)] + P(R \cap W) + P(W \cap C) + P(R \cap C)$$

Given $P(!(R \cup W \cup C)) = 1/6 \Rightarrow P(R \cup W \cup C) = 1 - 1/6 = 5/6$

substituting remaining values ~~we get~~ from above diagram:

$$P(R \cap W \cap C) = 5/6 - \left[\frac{1}{3} + \frac{1}{2} + \frac{1}{2}\right] + \left[\frac{1}{4} + \frac{1}{6} + \frac{1}{6}\right]$$

$$P(R \cap W \cap C) = \frac{5}{6} - \frac{2+3+3}{6} + \frac{3+2+2}{12}$$

$$= \frac{7}{12} + \frac{5}{6} - \frac{8}{6} = \frac{7}{12} - \frac{6}{12} = \frac{1}{12}$$

Thus $P(R \cap W \cap C) = P$ (Getting red, waterproof, cool toy)

$$= \frac{1}{12}$$

problem C :- Given $P(C) = 1/3$ and $P(R) = 1/2$

We have mentioned earlier $C = \{$ toys that are cool $\}$

then let $!C$ be the event that toys are not cool

$!C = \{$ toys that are not cool $\}$

we know from axioms of probability,

$$P(C) + P(!C) = 1$$

$$\text{Thus } P(!C) = 1 - \frac{1}{3} = \frac{2}{3}$$

The question is asking probability P (what is probability toy is not cool, given color is red)

this can be written as $P(!C|R)$

According to definition of conditional probability,

If $P(B) > 0$, $\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$

thus $P(!c|R) = \dfrac{P(!c\cap R)}{P(R)}$

From area diagram ~~and~~ $P(!c\cap R)$ = area under $\{e, h\}$

we know $P(c)+P(!c)=1$

$\Rightarrow P(!c) = 1 - P(c)$

$P(!c\cap R) = P((1-c)\cap R) = P(R) - P(R\cap C)$

$= \dfrac{1}{2} - \dfrac{1}{6}$

$P(!c\cap R) = \dfrac{3-1}{6} = \dfrac{2}{6} = \dfrac{1}{3}$

Substituting this back above,

$P(!c|R) = \dfrac{P(!c\cap R)}{P(R)} = \dfrac{1/3}{1/2} = \dfrac{2}{3} = 0.66$

Thus $P(!c|R) = P$ (toy is not cool, given color is red)

$= 2/3 = 0.66$

Problem (d):-

Find P (toy is cool, given randomly selected toy is red or waterproof)

$= P(c\,|\,R\cup W)$

Applying rule of conditional probability

$$\boxed{P(c|R\cup W) = \frac{P(c\cap(R\cup W))}{P(R\cup W)}} \quad \text{—— (5)}$$

we know from axioms of probability

$P(C \cap (R \cup W))$ can be written as

P Area under $\{i, k, j\}$

~~Pdf~~ · $P(C \cap (R \cup W))$ can be written as;

$$P(C \cup (R \cup W)) = P(C) + P(R \cup W) - P(C \cap (R \cup W))$$

$$\Rightarrow P(C \cap (R \cup W)) = P(C) + P(R \cup W) - P(C \cup (R \cup W))$$

we know $P(R \cup W \cup C) = 5/6$ and $P(C) = \frac{1}{3}$

$$P(R \cup W) = P(R) + P(W) - P(R \cap W) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} \quad \text{—— (q)}$$

Substituting,

$$P(C \cap (R \cup W)) = \frac{1}{3} + \frac{3}{4} - \frac{5}{6} = \frac{13}{12} - \frac{10}{12} = \frac{1}{4} \quad \text{—— (r)}$$

$$P(C \cap (R \cup W)) = 1/4$$

Substituting ~~or~~ back (r) and (q) in equation (s)

$$P(C \mid R \cup W) = \frac{P(C \cap (R \cup W))}{P(R \cup W)} = \frac{1/4}{3/4} = \frac{1}{3} = 0.33$$

Thus P (that a toy is tool, given a randomly selected toy is R or W) $= 0.33$

3. On the Overlap of Two Events

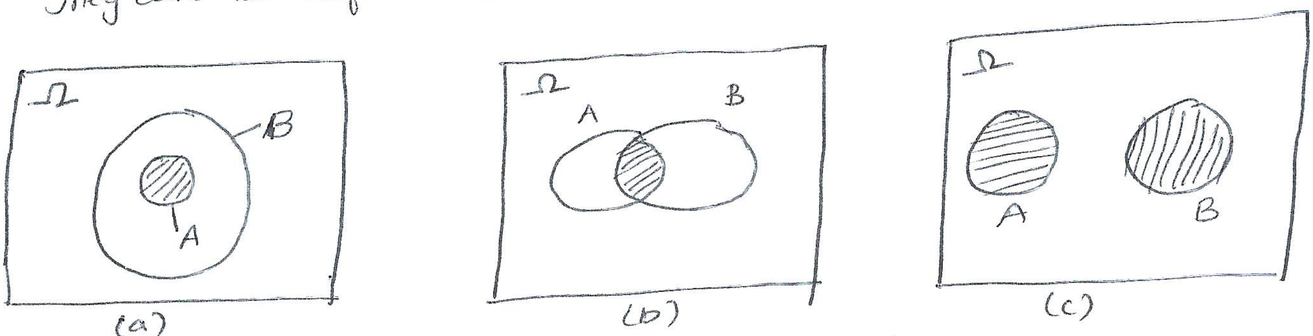
Suppose for events A and B, P(A) = 1/2, P(B) = 2/3, but we have no more information about the events.

(a) What are the maximum and minimum possible values for P (A ∩ B)? (3 points)

(b) What are the maximum and minimum possible values for P(A|B)? (3 points)

In order to solve this problem, we need to consider all the possible cases A and B can be spread in an event space 'Ω':

They can be defined as below:



Given ~~P(A∩B)~~ $P(A) = \frac{1}{2}$ and $P(B) = 2/3$

From the rules of probability $P(\Omega) = 1$ (the event space should ~~equal~~ equal to 1)

Case C:-

But in case (c), from axioms of probability

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Here $P(A \cap B) = 0$, as the two sets are not overlapping,

hence $P(A) + P(B) = P(A \cup B) = \frac{1}{2} + \frac{2}{3} \geq \frac{7}{6} > 1$

Thus, the two sets A and B in case (c) are not valid representation, as $P(\Omega) = 1$ and $P(A \cup B) \in [0, 1]$

but $P(A \cup B) = 7/6 > 1$

Hence we can ignore case (c).

Case A:-

In case (a), A is a subset of B. The ~~[illegible]~~ reverse of it is not possible as $P(A) = \frac{1}{2} < P(B) = \frac{2}{3}$

In case (a), from diagram it is imperative that

$P(A \cap B) = \frac{1}{2}$ since, A ~~P(A)~~ is a subset of B

thus $\boxed{P(A \cap B) = P(A) = \frac{1}{2} = 0.5}$ —— (d)

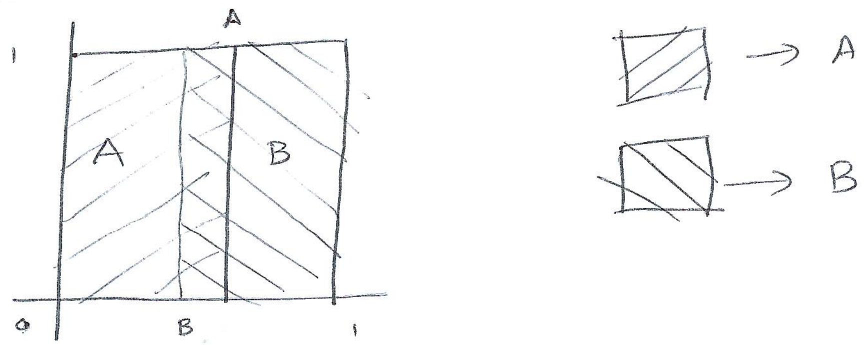
Now, we can derive $P\left(\frac{A}{B}\right)$ using the ~~axiom~~ rule of conditional probability,

$$P\left(\frac{A}{B}\right) = \frac{P(A \cap B)}{P(B)}$$

substituting $P(B) = \frac{2}{3}$ and $P(A \cap B) = 0.5$ from (d) above,

$$\boxed{P\left(\frac{A}{B}\right) = \frac{1/2}{2/3} = \frac{3}{4} = 0.75} \quad \text{—— (e)}$$

Case b:- For case (b) we can represent space as



From axioms of probability we know,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\Rightarrow P(A \cap B) = P(A) + P(B) - P(A \cup B)$$

substituting $P(A)$, $P(B)$ and $P(A \cup B) = 1$ (since A and B cover entire event space)

$$P(A \cap B) = \frac{1}{2} + \frac{2}{3} - 1 = \frac{7}{6} - 1 = \frac{1}{6}$$

Thus $\boxed{P(A \cap B) = 1/6}$ —— (f)

Now, we can derive $P\left(\frac{A}{B}\right)$ as,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

$P(A \cap B) = 1/6$ and $P(B) = 2/3$

$$\Rightarrow P(A|B) = \frac{1/6}{2/3} = \frac{1}{6} \times \frac{3}{2} = \frac{1}{4}$$

$$\boxed{P(A|B) = 1/4} \quad \text{——— (h)}$$

Thus ~~from~~ from derived values in (d), (e), (f), (g)

we can find out

minimum $P(A \cap B) = 1/6 = 0.16$ and minimum $P(A|B) = 1/4 = 0.25$

which happens when A and B overlap and A and B occupy the entire event space

maximum $P(A \cap B) = 0.5$ and maximum $P(A|B) = 0.75$

which happens when A is a subset of B and $A \cap B = A$

~~These~~ In This case, A and B are disjoint events

4. Can't Please Everyone! Among Berkeley students who have completed w203, 3/4 like statistics. Among Berkeley students who have not completed w203, only 1/4 like statistics. Assume that only 1 out of 100 Berkeley students completes w203. Given that a Berkeley student likes statistics, what is the probability that they have completed w203? (3 points)

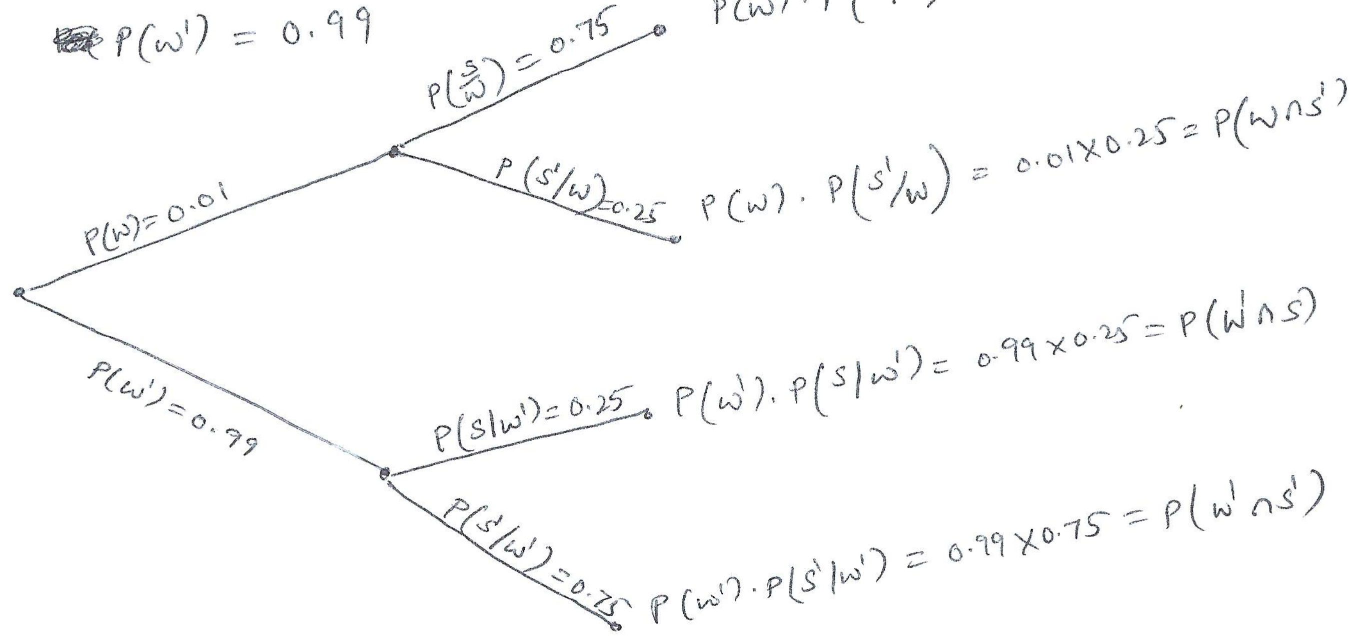
This can be interpreted in tree leaf as,

Event $W = \{$ student completed w203 $\}$ $\quad W' = \{$ students who didnot complete w203 $\}$

$S = \{$ Student likes statistics $\}$ $\quad S' = \{$ Student who does not like statistics $\}$

$P(W) = 1/100 = 0.01$

$P(W') = 0.99$



$P($ Given Berkeley student likes statistics, what is probability they have completed w203 $)$

$= P\left(\frac{W}{S}\right)$

According to Bayes' theorem

$$\boxed{P(A_j/B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B/A_j) \cdot P(A_j)}{\sum_{i=1}^{k} P(B/A_i) \cdot P(A_i)} \quad j = 1 \dots k}$$

thus $P\left(\frac{W}{S}\right) = \frac{P(W \cap S)}{P(S)} = \frac{P(S/W) \cdot P(W)}{P(S/W) \cdot P(W) + P\left(\frac{S}{W'}\right) \cdot P(W')}$

$$= \frac{0.01 \times 0.75}{0.01 \times 0.75 + 0.99 \times 0.25}$$

$$= \frac{0.0075}{0.0075 + 0.2475}$$

$$= \frac{0.0075}{0.255} = 0.0294$$