Name: Ivan Leung

Instructions

Please read the following instructions carefully:

1. Please show all notation for probability statements.
2. Box your final answers.
3. Please verify that your scans are legible.
4. Please assign pages the the questions when submitting to gradescope.
5. This assignment is due via gradescope on the due date.

1. For each of the following, state what the sample space is.

   (a) You select 1 card from a 52 card deck and observe the suit of the card.

   $\boxed{\{\text{diamond}, \text{shade}, \text{heart}, \text{club}\}}$

   (b) How many days a week someone goes to the gym.

   $\boxed{\{0, 1, 2, 3, 4, 5, 6, 7\}}$

   (c) What size drink someone orders from Starbucks (short, tall, grande, venti, trenta).

   $\boxed{\{\text{short}, \text{tall}, \text{grande}, \text{venti}, \text{trenta}\}}$

   (d) The sum of the roll of 2 dice

   $\boxed{\{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}}$

2. Suppose the $P(A) = 0.3$ and $P(B) = 0.7$.

(a) Can you compute P(A and B) if you only know P(A) and P(B)?

No, we don't know if A and B are independent, so we can't compute P(A and B).

(b) Assuming that events A and B arise from independent random processes:

i. (points) what is P(A and B)?

$$P(A \text{ and } B) = P(A)P(B) = 0.3(0.7)$$
$$= \boxed{0.21}$$

ii. (points) what is P(A or B)?

$$P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$$
$$= 0.3 + 0.7 - 0.21$$
$$= \boxed{0.79}$$

iii. (points) what is $P(A|B)$?

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.21}{0.7} = \boxed{0.3}$$

(c) If we are given that P(A and B) = 0.1, are the random variables giving rise to events A and B being independent events in this case?

No, since the product P(A) and P(B) is equal to 0.21, they are dependent events.

(d) If we are given that P(A and B) = 0.1, what is $P(A|B)$?

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)} = \frac{0.1}{0.7} = \boxed{0.1429}$$

3. It's never lupus. Lupus is a medical phenomenon where antibodies that are supposed to attack foreign cells to prevent infections instead see plasma proteins as foreign bodies, leading to a high risk of blood clotting. It is believed that 2% of the population suffer from this disease. The test is 98% accurate if a person actually has the disease. The test is 74% accurate if a person does not have the disease.

   (a) Draw a tree diagram of the scenario described above including all appropriate notation.

$P(d)$ 0.02

$\rightarrow P(p \mid d)$ 0.98 $\longrightarrow P(p \cap d) = (0.02)(0.98) = 0.0196$

$\rightarrow P(p^c \mid d)$ 0.02 $\longrightarrow P(p^c \cap d) = (0.02)(0.02) = 0.0004$

$P(d^c)$ 0.98

$\rightarrow P(p^c \mid d^c)$ 0.74 $\longrightarrow P(p^c \cap d^c) = (0.98)(0.74) = 0.7252$

$\rightarrow P(p \mid d^c)$ 0.26 $\longrightarrow P(p \cap d^c) = (0.98)(0.26) = 0.2548$

   (b) What is the probability that a person has lupus and tests positive for the disease?

$P(p \cap d) = \boxed{0.0196}$

   (c) What is the probability that a person tests positive for the disease?

$P(p \cap d) + P(p \cap d^c) = 0.0196 + 0.2548$

$= \boxed{0.2744}$

   (d) There is a line from the Fox television show House that is often used after a patient tests positive for lupus: 'It's never lupus.' Do you think there is truth to this statement?

It is somewhat true because the test's accuracy is low when someone does not have the disease.

4. A customer is randomly selected from a coffee shop. Let A be the event the selected customer ordered a beverage, and B be the event the customer ordered food. Suppose $P(A) = 0.65$, $P(B) = 0.45$, and $P(A \cup B) = 0.98$.

   (a) What is the probability that the customer did not order food nor a beverage?

   P(not order food nor a beverage) $= 1 - P(A \cup B)$

   $= 1 - 0.98$

   $= \boxed{0.02}$

   (b) What is the probability that the customer ordered food and a beverage?

   P(order food $\cap$ order a beverage)

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

   $= 0.65 + 0.45 - 0.98$

   $= \boxed{0.12}$

   (c) What is the probability the customer ordered a beverage but not food? (Hint: You will use the answer from part b.)

   P(order a beverage but not food)

   $= P(A \cap B^c) = P(A) - P(A \cap B)$

   $= 0.65 - 0.12 = \boxed{0.53}$

   (d) Are the events of ordering a beverage and ordering food mutually exclusive?

   No, $P(A) \cdot P(B) = 0.2925$, so they are not mutually exclusive.

5. Data collected at elementary schools in DeKalb County, GA suggest that each year roughly 25% of students miss exactly one day of school, 15% miss 2 days, and 28% miss 3 or more days due to sickness.

   (a) What is the probability that a student chosen at random does not miss any days of school due to sickness this year?

   $P(\text{not sick}) = 1 - 0.25 - 0.15 - 0.28$
   $= \boxed{0.32}$

   (b) What is the probability that a student chosen at random misses no more than one day?

   $P(\text{miss one day}) = P(0) + P(1)$
   $= 0.32 + 0.25 = \boxed{0.57}$

   (c) What is the probability that a student chosen at random misses at least one day?

   $P(\text{sick}) = 1 - P(\text{not sick}) = 1 - 0.32 = \boxed{0.68}$

   (d) If a parent has two kids at a DeKalb County elementary school, what is the probability that neither kid will miss any school? Note any assumption you must make to answer this question.

   $P(1^{st} \text{ kid not sick} \cap 2^{nd} \text{ kid not sick})$
   $= (0.32)(0.32) = \boxed{0.1024}$

   (e) If a parent has two kids at a DeKalb County elementary school, what is the probability that both kids will miss some school, i.e. at least one day? Note any assumption you make.

   $P(1^{st} \text{ kid sick} \cap 2^{nd} \text{ kid sick})$
   $= (0.68)(0.68) = \boxed{0.4624}$

   (f) If you made an assumption in part (d) or (e), do you think it was reasonable? If you didnât make any assumptions, double check your earlier answers.

   I assume that the events are independent.

6. A 2010 Pew Research poll asked 1,306 Americans “From what youâve read and heard, is there solid evidence that the average temperature on earth has been getting warmer over the past few decades, or not?”. The table below shows the distribution of responses by party and ideology, where the counts have been replaced with relative frequencies.

		Earth is warming	Not warming	Donât Know Refuse	Total
	Conservative Republican	0.11	0.20	0.02	0.33
Party and	Mod/Lib Republican	0.06	0.06	0.01	0.13
Ideology	Mod/Cons Democrat	0.25	0.07	0.02	0.34
	Liberal Democrat	0.18	0.01	0.01	0.20
	Total	0.60	0.34	0.06	1.00

(a) Are believing that the earth is warming and being a liberal Democrat mutually exclusive?

$P(\text{earth is warming} \cap \text{liberal dem}) = 0.18 > 0$

therefore the two events are $\boxed{\text{not mutually exclusive}}$

(b) What is the probability that a randomly chosen respondent believes the earth is warming or is a liberal Democrat?

$P(\text{earth is warming} \cup \text{liberal dem}) = P(\text{warm}) + P(\text{lib dem}) - P(\text{warm} \cap \text{lib dem})$

$= 0.60 + 0.20 - 0.18 = \boxed{0.62}$

(c) What is the probability that a randomly chosen respondent believes the earth is warming given that he is a liberal Democrat?

$P(\text{warm} \mid \text{lib dem}) = \dfrac{0.18}{0.20} = \boxed{0.90}$

(d) What is the probability that a randomly chosen respondent believes the earth is warming given that he is a conservative Republican?

$P(\text{warm} \mid \text{co Rep}) = \dfrac{0.11}{0.33} = \boxed{0.3333}$

(e) Does it appear that whether or not a respondent believes the earth is warming is independent of their party and ideology? Explain your reasoning.

$P(\text{warm} \mid \text{lib dem}) \stackrel{?}{=} P(\text{warm})$

$0.90 \neq 0.60$

therefore, they are $\boxed{\text{dependent}}$

(f) What is the probability that a randomly chosen respondent is a moderate/liberal Republican given that he does not believe that the earth is warming?

$P(\text{mod/lib Rep} \mid \text{warm}) = \dfrac{0.06}{0.34} = \boxed{0.1765}$

7. This problem will use R. The following code will simulate flipping a coin 100 times.

```
n <- 100
space <- c("H","T")
p <- c(0.5,0.5)
observation <- sample(space, size=n, prob=p, replace=TRUE)
p.hat <- sum(observation=="H")/n
p.hat
```

(a) Run the code 5 times, and report the proportion of times a head showed up for each time you ran the code. Label them $\hat{p}_1, \hat{p}_2, \hat{p}_3, \hat{p}_4, \hat{p}_5$,

$\hat{p}_1 = 0.5$ $\hat{p}_4 = 0.46$
$\hat{p}_2 = 0.44$ $\hat{p}_5 = 0.46$
$\hat{p}_3 = 0.52$

(b) Now change n to 10000, n $<-$ 10000. Run the code 5 times, and report the proportion of times a head showed up for each time you ran the code. Label them $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$, $\hat{p}_4$,$\hat{p}_5$,

$\hat{p}_1 = 0.4991$ $\hat{p}_4 = 0.4949$
$\hat{p}_2 = 0.5003$ $\hat{p}_5 = 0.4976$
$\hat{p}_3 = 0.4961$

(c) Noting that the true proportion should be 0.5, what do you notice as n increases.

As n increases, the proportion gets closer to the true proportion

(d) Now, set the seed to your computer to 55. Run the code 5 times report the proportion of times a head showed up for each time you ran the code.

```
set.seed(55)
n <- 100
space <- c("H","T")
p <- c(0.5,0.5)
observation <- sample(space, size=n, prob=p, replace=TRUE)
p.hat <- sum(observation=="H")/n
p.hat
```

$\hat{p}_1 = \hat{p}_2 = \hat{p}_3 = \hat{p}_4 = \hat{p}_5 = 0.53$

(e) Is your answer the same or different than in part (a)? What do you notice when you set the seed?

The answer is different than in part (a). When the seed is set, the result is always the same.