Problem Set - 3 (Probability Definitions, Joint and Conditional Probability)

ECO 104 (Section 8)
Instructor: Shaikh Tanvir Hossain

Due Date : 25th November, 2023, Saturday (submit in class), individual assignment

§

You should write the solutions neatly in paper and submit in class. **This is NOT a group assignment, this is an individual assignment**, so please do the problems and submit individually! Please ask me in Ed if you have any question.

§. Theory and Methods

1. From Anderson et al. (2020), Chapter 4, #30

   Suppose that we have two events, A and B, with $P(A) = .50, P(B) = .60$, and $P(A \cap B) = .40$

   (a) Find $P(A \mid B)$.

   (b) Find $P(B \mid A)$.

   (c) Are A and B independent? Why or why not?

   Solution:

   a. and b.

   Apply the definition of Conditional Probability. $P(A|B) = \frac{P(A \cap B)}{P(B)}$ and $P(B|A) = \frac{P(A \cap B)}{P(A)}$

   c. NO they are not, the reason is, by the definition of independence, two events A and B are independent if $P(A \cap B) = P(A) \times P(B)$. Since $P(A \cap B) = .40$ and $P(A) \times P(B) = .30$, this means the two sets are not independent.

2. From Anderson et al. (2020), Chapter 4, #31

   Assume that we have two events, A and B, that are mutually exclusive. Assume further that we know $P(A) = .30$ and $P(B) = .40$.

   (a) What is $P(A \cap B)$?

   (b) What is $P(A \mid B)$?

   (c) A student in statistics argues that the concepts of mutually exclusive events and independent events are really the same, and that if events are mutually exclusive they must be independent. Do you agree with this statement? Use the probability information in this problem to justify your answer.

   (d) What general conclusion would you make about mutually exclusive and independent events given the results of this problem?

   Solution: a. If two events are mutually exclusive, this means they are disjoint so $A \cap B = \emptyset$. In this case we know that $P(A \cap B) = P(\emptyset) = 0$.

   b. $P(A|B) = 0$ since $P(A \cap B) = 0$.

   c. NO, definitely they are not same. Mutually exclusive means they are disjoint so probability of their joint event is 0. On the other hand if the two events are independent, their probability may or may not be 0. For example in this case $P(A \cap B) = 0$, we know this from the question since it's given that they are disjoint. But if we calculate $P(A) \times P(B)$, then we get $.12 \neq 0$. So clearly $P(A \cap B) \neq P(A) \times P(B)$. This means they are not independent here. So two sets are disjoint does not automatically mean they are independent. Independence is a different concept. It's defined in a way such that the probability of the joint event is same as the product of the marginal probabilities.

3. From Anderson et al. (2020), Chapter 4, # 39.

   The prior probabilities for events A_1 and A_2 are $P(A_1) = .40$ and $P(A_2) = .60$. It is also known that $P(A_1 \cap A_2) = 0$. Suppose $P(B \mid A_1) = .20$ and $P(B \mid A_2) = .05$.

(a) Are A_1 and A_2 mutually exclusive? Explain.

(b) Compute $P(A_1 \cap B)$ and $P(A_2 \cap B)$.

(c) Compute $P(B)$.

(d) Apply Bayes' theorem to compute $P(A_1 \mid B)$ and $P(A_2 \mid B)$.

Solution: a. We need to look at the options carefully. Here we have $P(A_1 \cap A_2) = 0$. Now it can be 0 in either of two cases, 1) $A_1 \cap A_2 = \emptyset$ or 2) A_1 and A_2 are independent and $P(A_1) = 0$ and $P(A_2) = 0$. The second option is not possible here since $P(A_1) = .40$ and $P(A_2) = .60$ and $P(A_1 \cap A_2) = 0$, so they are definitely not independent. This means they are disjoint or mutually exclusive.

b. Apply the Multiplication rule of Conditional Probability.

c. Applying Law of Total Probability gives us

$$P(B) = P(B|A_1)P(A_1) + P(B|A_2)P(A_2)$$

Now you can calculate $P(B)$

d. With the help of c., and applying Bayes Theorem, we can calculate $P(A_1|B)$ and $P(A_2|B)$. We will do one here, please do the other,

$$\begin{aligned} P(A_1|B) &= \frac{P(A_1 \cap B)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B)} \\ &= \frac{P(B|A_1)P(A_1)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2)} \end{aligned}$$

Just plug the probabilities from the question.

4. From Anderson et al. (2020), Chapter 4, #40.

The prior probabilities for events A_1, A_2, and A_3 are $P(A_1) = .20, P(A_2) = .50$, and $P(A_3) = .30$. The conditional probabilities of event B given A_1, A_2, and A_3 are $P(B \mid A_1) = .50, P(B \mid A_2) = .40$, and $P(B \mid A_3) = .30$.

(a) Compute $P(B \cap A_1), P(B \cap A_2)$, and $P(B \cap A_3)$.

(b) Apply Bayes' theorem, equation, to compute the posterior probability $P(A_2 \mid B)$.

(c) Use the tabular approach to applying Bayes' theorem to compute $P(A_1 \mid B)$, $P(A_2 \mid B)$, and $P(A_3 \mid B)$.

Solution:

a. Just apply the Multiplication rule of Conditional Probability.

b. The last problem was similar

$$\begin{aligned} P(A_2|B) &= \frac{P(A_2 \cap B)}{P(B)} \\ &= \frac{P(B|A_2)P(A_2)}{P(B)} \\ &= \frac{P(B|A_2)P(A_2)}{P(B|A_1)P(A_1) + P(B|A_2)P(A_2) + P(B|A_3)P(A_3)} \end{aligned}$$

c. Applying the similar formula you can calculate $P(A_1|B)$. Then once you have $P(A_1|B)$ and $P(A_2|B)$, you can calculate $P(A_3|B) = 1 - P(A_1|B) - P(A_2|B)$. This is because Conditional Probability function acts like a Probability function, so the Conditionals Probabilities should be summed to 1 for the events which are actually a partition.

§. **Applied Problems**

5. From Anderson et al. (2020), Chapter 4, #25

   A 2018 Pew Research Center survey (Pew Research website) examined the use of social media platforms in the United States. The survey found that there is a $.68$ probability that a randomly selected American will use Facebook and a $.25$ probability that a randomly selected American will use LinkedIn. In addition, there is a $.22$ probability that a randomly selected American will use both Facebook and LinkedIn.

   (a) What is the probability that a randomly selected American will use Facebook or LinkedIn?

   (b) What is the probability that a randomly selected American will not use either social media platform?

   Solution:

   Let's denote the event that an American uses Facebook by F, and an American uses LinkedIn by L. Essentially this is the set of people who uses Facebook and LinkedIn respectively. Question tells us $P(F) = .68$, $P(L) = .25$ and $P(F \cap L) = .22$.

   a. This is easy, you need to calculate $P(F \cup L)$, just apply Question 1 (b)

   b. This asks to calculate $P((F \cup L)^c)$. You can calculate this in two ways. Using the answer from a. and then you can do $1-$ the answer from a. Or you can also apply DeMorgan's Law.

6. From Anderson et al. (2020), Chapter 4, #27.

   A marketing firm would like to test-market the name of a new energy drink targeted at 18- to 29-year-olds via social media. A study by the Pew Research Center found that 35% of U.S. adults (18 and older) do not use social media (Pew Research Center website, October 2015). The percentage of U.S. young adults age 30 and older is 78%. Suppose that the percentage of the U.S. adult population that is either age 18-29 or uses social media is 67.2%.

   (a) What is the probability that a randomly selected U.S. adult uses social media?

   (b) What is the probability that a randomly selected U.S. adult is aged 18-29?

   (c) What is the probability that a randomly selected U.S. adult is 18-29 and a user of social media?

   Solution: Let A be the event that an adult uses social media and B be the event that an adult is aged 18-29.

   a. Form the question we know $P(A^c) = .35$, so $P(A) = .65$

   b. From the question we know $P(B^c) = .78$, this means $P(B) = .22$

   c. From the question we know $P(A \cup B) = .672$. So with this we can calculate $P(A \cap B)$.

7. From Anderson et al. (2020), Chapter 4, #29.

   High school seniors with strong academic records apply to the nation's most selective colleges in greater numbers each year. Because the number of slots remains relatively stable, some colleges reject more early applicants. Suppose that for a recent admissions class, an Ivy League college received 2851 applications for early admission. Of this group, it admitted 1033 students early, rejected 854 outright, and deferred 964 to the regular admission pool for further consideration. In the past, this school has admitted 18% of the deferred early admission applicants during the regular admission process. Counting the students admitted early and the students admitted during the regular admission process, the total class size was 2375 .

- Let E be the event that a student who applies for early admission is admitted early
- R be the event that a student who applies for early admission is rejected outright
- and D be the event that a student who applies for early admission is deferred to the regular admissions pool.

(a) Use the data to estimate $P(E), P(R)$, and $P(D)$.

(b) Are events E and D mutually exclusive? Find $P(E \cap D)$.

(c) For the 2375 students who were admitted, what is the probability that a randomly selected student was accepted during early admission?

(d) Suppose a student applies for early admission. What is the probability that the student will be admitted for early admission or be deferred and later admitted during the regular admission process?

Solution: a.

$$P(E) = 1033/2851$$
$$P(R) = 854/2851$$
$$P(D) = 964/2851$$

b. Yes they are mutually exclusive because a student cannot be simultaneously admitted early and then also deferred. So this means $P(E \cap D) = 0$.

c. Not sure how to do this, will give you answer later.....

d. This is simply $P(E \cup D) = P(E) + P(D)$. We can add here because the events are disjoint.

8. From Anderson et al. (2020), Chapter 4, #32.

Consider the following example survey results of 18 - to 34 -year olds in the United States, in response to the question "Are you currently living with your family?"

	Yes	No	Totals
Men	106	141	247
Women	92	161	253
Totals	198	302	500

(a) Develop the joint probability table for these data and use it to answer the following questions.

(b) What are the marginal probabilities?

(c) What is the probability of living with family given you are an 18-to 34-year-old man in the United States?

(d) What is the probability of living with family given you are an 18-to 34-year-old woman in the United States?

(e) What is the probability of an 18-to 34-year-old in the United States living with family?

(f) If, in the United States, 49.4% of 18- to 34-year-olds are male, do you consider this a good representative sample? Why?

Solution:

a) Use your calculator to calculate joint probabilities,

	Yes	No	Totals
Men	.212	.282	.494
Women	.184	.322	.506
Totals	.396	.604	1

b) The marginal probabilities are

$$P(\text{Men}) = .494, P(\text{Women}) = .506, P(\text{Yes}) = .396, P(\text{No}) = .604.$$

c) This is asking $P(\text{Yes}|\text{Men})$. You can directly calculate this from the joint probability table

$$P(\text{Yes}|\text{Men}) = \frac{P(\text{Men} \cap \text{Yes})}{P(\text{Men})} = \frac{.212}{.494}$$

Or you can also use the joint frequency table to do this calculation, this will be simply $\frac{106}{247}$. This is going to be the same number if we do $\frac{.212}{.494}$.

d) Similar to c)

e) This is $P(\text{Yes}) = .396$

f) Yes! It looks like my sample has 49.4% male, this is exactly same as my population. So this sample is a good representative of the population.

9. From Anderson et al. (2020), Chapter 4, #33.

Students taking the Graduate Management Admissions Test (GMAT) were asked about their undergraduate major and intent to pursue their MBA as a full-time or part-time student. A summary of their responses follows.

		Majors			
		Business	Engineering	Other	Totals
Intended	Full-Time	352	197	251	800
Enrollment	Part-Time	150	161	194	505
Status	Totals	502	358	445	1305

(a) Develop a joint probability table for these data.

(b) Use the marginal probabilities of undergraduate major (business, engineering, or other) to comment on which undergraduate major produces the most potential MBA students.

(c) If a student intends to attend classes full-time in pursuit of an MBA degree, what is the probability that the student was an undergraduate engineering major?

(d) If a student was an undergraduate business major, what is the probability that the student intends to attend classes full-time in pursuit of an MBA degree?

(e) Let A denote the event that the student intends to attend classes full-time in pursuit of an MBA degree, and let B denote the event that the student was an undergraduate business major. Are events A and B independent? Justify your answer.

Solution: a. Very similar to the last problem, so use your calculator and finish the calculation. For example, $P(B \cap A) = \frac{352}{1305}$

		Majors			
		Business (B)	Engineering (E)	Other (O)	Totals
Intended	Full-Time (A)	$P(B \cap A)$	$P(E \cap A)$	$P(O \cap A)$	$P(A)$
Enrollment	Part-Time (Pt)	$P(B \cap Pt)$	$P(E \cap Pt)$	$P(O \cap Pt)$	$P(Pt)$
Status	Totals	$P(B)$	$P(E)$	$P(O)$	1

b. Figure out calculating $P(B), P(E)$ and $P(O)$, which is the highest.

c. It is asking to calculate $P(E|A)$, you can easily calculate this using the joint probability table.

d. It is asking to calculate $P(A|B)$, you can easily calculate this using the joint probability table.

e. In this case we need to check $P(B \cap A) = P(A) \times P(B)$, if this is not the case, then they are not independent.

10. From Anderson et al. (2020), Chapter 4, #37.

A 2018 Pew Research Center survey found that more Americans believe they could give up their televisions than could give up their cell phones (Pew Research website). Assume that the following table represents the joint probabilities of Americans who could give up their television or cell phone.

		Give Up Television		
		Yes	No	
Could Give Up	Yes	.31	.17	.48
Cellphone	No	.38	.14	.52
		.69	.31	

(a) What is the probability that a person could give up her cell phone?

(b) What is the probability that a person who could give up her cell phone could also give up television?

(c) What is the probability that a person who could not give up her cell phone could give up television?

(d) Is the probability a person could give up television higher if the person could not give up a cell phone or if the person could give up a cell phone?

Solution: We have the following joint probability table.

- Let C denote the event if the person is willing to give up the cell phone, and NC denote the event if the person is NOT willing to give up the cell phone.
- Similarly let T denote the event if the person is willing to give up the television, and NT denote the event if the person is NOT willing to give up the television.

		Give Up Television		
		Yes (T)	No (NT)	
Could Give Up	Yes (C)	.31	.17	.48
Cellphone	No (NC)	.38	.14	.52
		.69	.31	

a. This asks $P(C) = .48$

b. This is $P(C \cap T) = .31$

c. $P(NC \cap T) = .38$

d. Here the question asks to compare two conditional probabilities, $P(T|C)$ v.s. $P(T|NC)$.

11. From Anderson et al. (2020), Chapter 4, #38.

The Institute for Higher Education Policy, a Washington, D.C.-based research firm, studied the payback of student loans for 1.8 million college students who had student loans that began to become due six years ago (The Wall Street Journal). The study found that 50% of the student loans were being paid back in a satisfactory fashion, whereas 50% of the student loans were delinquent. The following joint probability table shows the probabilities of the student loan status and whether or not the student had received a college degree.

		College Degree		
		Yes	No	
Loan Status	Satisfactory	$.26$	$.24$	$.50$
	Delinquent	$.16$	$.34$	$.50$
		$.42$	$.58$	

(a) What is the probability that a student with a student loan had received a college degree?

(b) What is the probability that a student with a student loan had not received a college degree?

(c) Given the student had received a college degree, what is the probability that the student has a delinquent loan?

(d) Given the student had not received a college degree, what is the probability that the student has a delinquent loan?

(e) What is the impact of dropping out of college without a degree for students who have a student loan?

Solution: a. $P(\text{Yes}) = .42$

b. $P(\text{No}) = .58$

c. $P(\text{Delinquent}|\text{Yes})$

d. $P(\text{Delinquent}|\text{No})$

e. No idea, need to think, will tell you answer later! Let me know if you know the answer.

12. From Anderson et al. (2020), Chapter 4, #43.

According to a 2018 article in Esquire magazine, approximately 70% of males over age 70 will develop cancerous cells in their prostate. Prostate cancer is second only to skin cancer as the most common form of cancer for males in the United States. One of the most common tests for the detection of prostate cancer is the prostate-specific antigen (PSA) test. However, this test is known to have a high false-positive rate (tests that come back positive for cancer when no cancer is present). Suppose there is a $.02$ probability that a male patient has prostate cancer before testing. The probability of a false-positive test is $.75$, and the probability of a false-negative (no indication of cancer when cancer is actually present) is $.20$.

(a) What is the probability that the male patient has prostate cancer if the PSA test comes back positive?

(b) What is the probability that the male patient has prostate cancer if the PSA test comes back negative?

(c) For older men, the prior probability of having cancer increases. Suppose that the prior probability of the male patient is $.3$ rather than $.02$. What is the probability that the male patient has prostate cancer if the PSA test comes back positive? What is the probability that the male patient has prostate cancer if the PSA test comes back negative?

(d) What can you infer about the PSA test from the results of parts (a), (b), and (c)?

Solution:

		Test		
		Pos	Neg	
Cancer	Yes	?	?	.02
	No	?	?	.98
		?	?	

Let's write down what we have

- $P(Yes) = .02$
- $P(No) = .98,$
- $P(Neg|Yes) = .20,$
- $P(Pos|No) = .75$
- a) This is asking to calculate $P(Yes|Pos)$, this can be done using the Bayes' Rule. First note we can calculate $P(Pos|Yes) = 1 - .20 = .80$

$$P(Yes|Pos) = \frac{P(Pos|Yes) \times P(Yes)}{P(Pos)} = \frac{P(Pos|Yes) \times P(Yes)}{P(Pos|Yes) \times P(Yes) + P(Pos|No) \times P(No)}$$

- b) This is asking to calculate $P(Yes|Neg)$. Again apply Bayes Rule. Again note we can calculate $P(Neg|No) = 1 - .75 = .25$

$$P(Yes|Neg) = \frac{P(Neg|Yes) \times P(Yes)}{P(Neg)} = \frac{P(Neg|Yes) \times P(Yes)}{P(Neg|Yes) \times P(Yes) + P(Neg|No) \times P(No)}$$

- c) This is again asking to calculate $P(Yes|Pos)$, but this time we have $P(Yes) = .3$. So we need to apply Bayes Rule again just adjust this information.
- I think what we need to look here is $P(Yes|Pos) > P(No|Pos)$, and $P(No|Neg) > P(Yes|Neg)$. Then the PSA test is good.

13. From Anderson et al. (2020), Chapter 4, #44.

ParFore created a website to market golf equipment and golf apparel. Management would like a special pop-up offer to appear for female website visitors and a different special pop-up offer to appear for male website visitors. From a sample of past website visitors, ParFore's management learned that 60% of the visitors are male and 40% are female.

(a) What is the probability that a current visitor to the website is female?

(b) Suppose 30% of ParFore's female visitors previously visited the Dillard's department store website and 10% of ParFore's male visitors previously visited the Dillard's department store website. If the current visitor to ParFore's website previously visited the Dillard's website, what is the revised probability that the current visitor is female? Should the ParFore's website display the special offer that appeals to female visitors or the special offer that appeals to male visitors?

Solution:

So let's denote the event of the male visitors with M and the event of the female visitors with F.

a) This is directly given in the question, this will be $P(F) = .40$

b) Now we have a new event, that is ParFore's visitors have previsouly visisted Dillard or Not. Let's denote this event with D, then D^c is the event that ParFore's visitors didn't visit Dillard. So we can think about following joint table

	M	F	
D	?	?	?
D^c	?	?	?
	.60	.40	

where we only have two marginals. But we have following information $P(D|F) = .30$ and $P(D|M) = .10$

With this we can immidiately calculate $P(D^c|F) = .70$ and $P(D^c|M) = .90$ (this is because conditioning acts like updating sample space)

Question asked us to calculate $P(F|D)$, we need to use Bayes' theorem here

$$P(F|D) = \frac{P(D|F)P(F)}{P(D)} = \frac{P(D|F)P(F)}{P(D|F)P(F) + P(D|M)P(M)}$$

Now we can calculate this using the information that we have.

To answer the last question, we need to calculate $P(F|D)$ and $P(M|D)$, and check which is higher. If $P(M|D) > P(F|D)$, then ParFore should offer special appeals to male, if $P(F|D) > P(M|D)$ then ParFore should offer special appeals to female.

14. From Anderson et al. (2020), Chapter 4, #45.

The National Center for Health Statistics, housed within the U.S. Centers for Disease Control and Prevention (CDC), tracks the number of adults in the United States who have health insurance. According to this agency, the uninsured rates for Americans in 2018 are as follows: 5.1% of those under the age of $18, 12.4\%$ of those ages 18-64, and 1.1% of those 65 and older do not have health insurance (CDC website). Approximately 22.8% of Americans are under age 18 , and 61.4% of Americans are ages $18 - 64$.

(a) What is the probability that a randomly selected person in the United States is 65 or older?

(b) Given that the person is an uninsured American, what is the probability that the person is 65 or older?

Solution: Please try!

Remarks: Many problems are taken from Anderson et al. (2020). If possible you should do more problems from there.

References:

Anderson, D. R., Sweeney, D. J., Williams, T. A., Camm, J. D., Cochran, J. J., Fry, M. J. and Ohlmann, J. W. (2020), *Statistics for Business & Economics*, 14th edn, Cengage, Boston, MA.