Conditional Probability – Solutions

COR1-GB.1305 – Statistics and Data Analysis

Conditional Probability

1. Here is a table tabulated frequencies for the industries and genders of the students in the class survey.

	Gender		
Industry	Female	Male	Total
Finance	3	9	12
Manufacturing	2	4	6
Other	8	10	18
Total	13	23	36

(a) Express the following statements as conditional probabilities:

- $\frac{12}{32} \approx 38\%$ of the people working in Finance are Females.
- $\frac{3}{13} \approx 23\%$ of the Females work in Finance.

Solution:

$$P(\text{Female} \mid \text{Finance}) = \frac{12}{32},$$
$$P(\text{Finance} \mid \text{Female}) = \frac{3}{13}.$$

(b) Compute $P(\text{Manufacturing} \mid \text{Male})$ and $P(\text{Male} \mid \text{Manufacturing})$. Explain the difference between these two quantities.

Solution:

$$P(\text{Manufacturing} \mid \text{Male}) = \frac{4}{23},$$
$$P(\text{Male} \mid \text{Manufacturing}) = \frac{4}{6}.$$

The quantity $P(\text{Manufacturing} \mid \text{Male})$ is the proportion of Males who work in Manufacturing; the quantity $P(\text{Male} \mid \text{Manufacturing})$ is the proportion of people who work in Manufacturing that are Male.

2. The following table lists the pick-up and drop-off locations of approximately 170 million yellow cab taxi trips made in New York City in 2013. Numbers are reported in thousands.

Pick-up	Drop-off Bronx	Brooklyn	Manhattan	Queens	Staten Is.	Total
Bronx	53	1	37	4	0	95
Brooklyn	8	2,707	1,598	273	2	4,588
Manhattan	638	5,458	143,656	5,906	22	155,680
Queens	122	1,022	5,058	2,281	8	8,491
Staten Is.	0	0	0	0	3	3
Total	821	9,188	150,349	8,464	35	168,857

(a) Find $P(\text{drop-off Brooklyn} \mid \text{pick-up Manhattan})$ and $P(\text{pick-up Manhattan} \mid \text{drop-off Brooklyn})$. Explain the difference between these two quantities.

Solution:

$$P(\text{drop-off Brooklyn} \mid \text{pick-up Manhattan}) = \frac{5458}{155680} \approx 3.5\%,$$

$$P(\text{pick-up Manhattan} \mid \text{drop-off Brooklyn}) = \frac{5458}{9188} \approx 59.4\%.$$

3.5% of the rides that pick up in Manhattan drop off in Brooklyn; 59.4% of the rides that drop off in Brooklyn originate in Manhattan.

(b) Express the following statement as a conditional probability: "29% of the trips with drop-off locations in Brooklyn originated in the same borough."

Solution:

$$P(\text{pick-up Brooklyn} \mid \text{drop-off Brooklyn}) = \frac{2707}{9188} = 29\%.$$

Note:

$$P(\text{drop-off Brooklyn} \mid \text{pick-up Brooklyn}) = \frac{2707}{4588} = 59\%.$$

The Multiplicative Rule

3. Out of the 45 students enrolled in the class, 14 are female (31%) and 31 are male (69%). Suppose that we randomly select two different students.

   (a) What is the probability that both students are male?

   Solution: Define the two events

   $$\begin{aligned} A &= \text{the first student picked is male} \\ B &= \text{the second student picked is male.} \end{aligned}$$

   Then, $P(A) = \frac{31}{45}$, and $P(B \mid A) = \frac{30}{44}$. Thus, the probability that both will be male is

   $$\begin{aligned} P(A \cap B) &= P(A)P(B \mid A) \\ &= \frac{31}{45} \cdot \frac{30}{44} \\ &= \frac{930}{1980} \\ &= 47\%. \end{aligned}$$

   (b) What is the probability that both students are female?

   Solution: Using the events A and B defined in the previous part, $P(A^c) = \frac{14}{45}$ and $P(B^c \mid A^c) = \frac{13}{44}$. Thus, the probability that both will be female is

   $$\begin{aligned} P(A^c \cap B^c) &= P(A^c)P(B^c \mid A^c) \\ &= \frac{14}{45} \cdot \frac{13}{44} \\ &= \frac{182}{1980} \\ &= 9\%. \end{aligned}$$

   (c) What is the probability that one of the students is male and one of the students is female?

   Solution: The event "one student is male and the other is female" is equivalent to the compound event $(A \cap B^c) \cup (A^c \cap B)$; that is, either the first is male and the second is female, or the first is female and the second is male. Since $A \cap B^c$ and $A^c \cap B$ are mutually exclusive, it follows that

   $$P(\text{one male and one female}) = P(A \cap B^c) + P(A^c \cap B).$$

Using the multiplicative rule,

$$\begin{aligned} P(A \cap B^c) &= P(A)P(B^c \mid A) \\ &= \frac{31}{45}\frac{14}{44} \\ &= \frac{434}{1980} \end{aligned}$$

$$\begin{aligned} P(A^c \cap B) &= P(A^c)P(B \mid A^c) \\ &= \frac{14}{45}\frac{31}{44} \\ &= \frac{434}{1980}. \end{aligned}$$

Thus,

$$\begin{aligned} P(\text{one male and one female}) &= \frac{434}{1980} + \frac{434}{1980} \\ &= \frac{868}{1980} \\ &= 44\%. \end{aligned}$$

4. Of the 37 students who filled out the survey, 11 indicated that they drink at least one cup of coffee per day, while 26 indicated that they do not drink coffee on a typical day. Suppose that we randomly select two different survey respondents.

   (a) What is the probability that both students regularly drink coffee?

   Solution:
   $$\frac{11}{37} \cdot \frac{10}{36} = \frac{110}{1332} = 8\%.$$

   (b) What is the probability that neither student regularly drinks coffee?

   Solution:
   $$\frac{26}{37} \cdot \frac{25}{36} = \frac{650}{1332} = 49\%.$$

   (c) What is the probability that exactly one student regularly drinks coffee?

   Solution:
   $$\frac{11}{37} \cdot \frac{26}{36} + \frac{26}{37}\frac{11}{36} = \frac{572}{1332} = 43\%.$$

Independence

5. Suppose that you flip two fair coins. Let $A =$"the first coin shows Heads," $B =$ "The second coin shows Heads." Find the probability of getting Heads on both coins, i.e. find $P(A \cap B)$.

Solution: The long way to solve this problem is to write out the elementary outcomes and their probabilities:

Outcome	Probability
HH	$\frac{1}{4}$
HT	$\frac{1}{4}$
TH	$\frac{1}{4}$
TT	$\frac{1}{4}$

Since $A \cap B = \{\text{HH}\}$, it follows that

$$\mathrm{P}(A \cap B) = \tfrac{1}{4}.$$

We can solve this problem much more expediently using the independence of A and B:

$$\begin{aligned}\mathrm{P}(A \cap B) &= \mathrm{P}(A)\,\mathrm{P}(B \mid A)\\ &= \mathrm{P}(A)\,\mathrm{P}(B)\\ &= (\tfrac{1}{2})(\tfrac{1}{2})\\ &= \tfrac{1}{4}.\end{aligned}$$

6. Suppose that you roll three dice. What is the probability of getting exactly one 6?

Solution: Define the following events:

$$\begin{aligned}A &= \text{“6 on the first roll,”}\\ B &= \text{“6 on the second roll,”}\\ C &= \text{“6 on the third roll.”}\end{aligned}$$

Using the shorthand $\bar{A} = A^c$ and $A\bar{B}\bar{C} = A \cap \bar{B} \cap \bar{C}$, the event "exactly one 6" can be written as

$$\text{“exactly one 6”} = A\bar{B}\bar{C} \cup \bar{A}B\bar{C} \cup \bar{A}\bar{B}C.$$

These events are mutually independent, so

$$\mathrm{P}(\text{exactly one 6}) = \mathrm{P}(A\bar{B}\bar{C}) + \mathrm{P}(\bar{A}B\bar{C}) + \mathrm{P}(\bar{A}\bar{B}C).$$

Using the independence of events A, B, and C, we get

$$\begin{aligned}
P(A\bar{B}\bar{C}) &= P(A)\,P(\bar{B})\,P(\bar{C}) = (\tfrac{1}{6})(\tfrac{5}{6})(\tfrac{5}{6}) \\
P(\bar{A}B\bar{C}) &= P(\bar{A})\,P(B)\,P(\bar{C}) = (\tfrac{5}{6})(\tfrac{1}{6})(\tfrac{5}{6}) \\
P(\bar{A}\bar{B}C) &= P(\bar{A})\,P(\bar{B})\,P(C) = (\tfrac{5}{6})(\tfrac{5}{6})(\tfrac{1}{6})
\end{aligned}$$

Note that these three expressions are all equal. Thus,

$$P(\text{exactly one } 6) = 3 \cdot \tfrac{1}{6} \cdot \tfrac{5}{6} \cdot \tfrac{5}{6} \approx .347.$$

7. Suppose that you sell fire insurance policies to three different buildings in Manhattan, located in different neighborhoods. You estimate that the buildings have the following chances of being damaged by fire in the next 10 years: 0.2%, 5%, and 1%. Assume that fire damages to the three buildings are independent events. Compute the probability that exactly one building gets damaged by fire in the next 10 years.

Solution:

8. Suppose you have a database of 300K reviews from 15K businesses and 70K users. In each of the following scenarios, you randomly sample 2 reviews. Define events A and B as

$$\begin{aligned}
A &= \text{the first review is 4 or 5 stars} \\
B &= \text{the second review is 4 or 5 stars}
\end{aligned}$$

In which sampling schemes are events A and B independent? Assume that all samples are random and unbiased. Explain your answers.

(a) You sample two distinct reviews from the entire dataset.

Solution: Dependent, but very weakly so. (If the reviews are sampled with replacement, then they are independent.)

(b) You randomly sample one business from the dataset, then sample two distinct reviews of the business.

Solution: Dependent. If the first review is high, then the restaurant is likely good, and so the second review is likely high as well.

(c) You randomly sample one user from the dataset, then sample two distinct reviews written by the user.

Solution: Dependent. The first review tells you about the user, and that in turn tells you about the second review.

Bayes' Rule

9. Every year in March there is a standardized exam for people who want to be licensed sheep herders. It happens that, with probability 0.45, a person will pass this exam. In the process of screening people, it turns out that among those who passed the exam, 60% had taken college courses in biology. It happens also that 30% of all those who take the exam had college courses in biology. Find the probability that a person with college courses in biology will pass the exam.

Solution: The information in the problem is

$$\begin{aligned} \mathrm{P}(\mathrm{Pass}) &= .45 \\ \mathrm{P}(\mathrm{Bio}) &= .30 \\ \mathrm{P}(\mathrm{Bio} \mid \mathrm{Pass}) &= .60 \end{aligned}$$

The problem is asking us to compute the quantity $\mathrm{P}(\mathrm{Pass} \mid \mathrm{Bio})$. Using Bayes' rule,

$$\begin{aligned} \mathrm{P}(\mathrm{Pass} \mid \mathrm{Bio}) &= \mathrm{P}(\mathrm{Bio} \mid \mathrm{Pass}) \cdot \frac{\mathrm{P}(\mathrm{Pass})}{\mathrm{P}(\mathrm{Bio})} \\ &= (.60) \cdot \frac{(.45)}{(.30)} \\ &= .90. \end{aligned}$$

That is, there is a 90% chance that a person with college courses in biology will pass the exam.

10. In 2013, $92\% = \frac{155680}{168857}$ of taxi trips originated in Manhattan, and $5.4\% = \frac{9188}{168857}$ terminated in Brooklyn. Of all taxi trips originating in Manhattan, $3.5\% = \frac{5458}{155680}$ terminated in Brooklyn. What proportion of taxi trips with drop-off locations in Brooklyn originated in Manhattan?

Solution: We are given that

$$\begin{aligned} P(\text{pick-up Manhattan}) &= \frac{155680}{168857}, \\ P(\text{drop-off Brooklyn}) &= \frac{9188}{168857}, \\ P(\text{drop-off Brooklyn} \mid \text{pick-up Manhattan}) &= \frac{5458}{155680}. \end{aligned}$$

By Bayes' Rule,

$$\begin{aligned}
P(\text{pick-up Manhattan} \mid \text{drop-off Brooklyn}) &= P(\text{drop-off Brooklyn} \mid \text{pick-up Manhattan}) \\
&\quad \cdot \frac{P(\text{pick-up Manhattan})}{P(\text{drop-off Brooklyn})} \\
&= \frac{5458}{155680} \cdot \frac{155680/168857}{9188/168857} \\
&= \frac{5458}{9188} \\
&\approx 59.4\%.
\end{aligned}$$

11. **Lie Detector.** Through accounting procedures, it is known that 10% of the employees of a store are stealing. To find out who is stealing, the manager decides to make all employees to take a lie detector test. The lie detector is accurate 80% of the time: if an employee is a thief, then he or she will fail the test with probability 0.8; if an employee is honest, then he or she will pass the test with probability 0.8. In this activity we will simulate the results of the manager's investigation.

    (a) Use your smartphone (or your neighbor's smartphone) to go to `http://random.org`. Click the "Generate" button to draw a random number between 1 and 100. Write down your number. Everyone in the class should generate his or her own number. If your number is in the range 1–10, write "Thief"; if your number is in the range 11–100, write "Honest".

    (b) Click "Generate" again to generate a new random number. Write down the number. If the number is in the range 1–80, then lie detector gives the correct answer ("Fail" for thief, "Pass" for honest). If the number is in the range 81–100, then the lie detector gives the wrong answer and records "Pass" for thief and "Fail" for honest. Write down the result of the test.

    (c) What proportion of the people who failed the lie detector test are thieves?

    Solution: In class, we computed the probability that someone who fails the test is a thief. The problem tells us that P(Thief) = 0.1 and P(Fail | Thief) = 0.8. To use Bayes' rule, we need to compute P(Fail). This is given by:

    $$\mathrm{P(Fail)} = (0.1)(0.8) + (0.9)(0.2) = 0.26;$$

    to compute this probability, we have noted that to Fail, an employee is either a Thief and they Fail, or they are Honest and they Fail. Thus,

$$\begin{aligned}
\mathrm{P(Thief \mid Fail)} &= \mathrm{P(Fail \mid Thief)}\frac{\mathrm{P(Thief)}}{\mathrm{P(Fail)}} \\
&= (0.80)\frac{(0.10)}{(0.26)} \\
&= 0.31.
\end{aligned}$$

Thus, approximately 31% of the people who fail the lie detector test are thieves.