Conditional Probability – Solutions
COR1-GB.1305 – Statistics and Data Analysis

Conditional Probability

1. Here is a table of the tabulated frequencies for the expected starting salary and gender for the respondents to the class survey.

	Gender		
Salary ($1K)	Female	Male	Total
$(0, 100]$	11	4	15
$(100, 125]$	5	7	12
$(125, \infty]$	5	15	20
Total	21	26	47

(a) Express the following statements as conditional probabilities:

- $\frac{11}{21} \approx 52\%$ of the females listed a starting salary of $100K or lower.
- $\frac{11}{15} \approx 73\%$ of those listing starting salaries of $100K or lower are female.

Solution:

$$P(\text{Salary} \leq \$100\text{K} \mid \text{Female}) = \frac{11}{21},$$
$$P(\text{Female} \mid \text{Salary} \leq \$100\text{K}) = \frac{11}{15}.$$

(b) Compute $P(\text{Male} \mid \text{Salary} > \$125\text{K})$ and $P(\text{Salary} > \$125\text{K} \mid \text{Male})$. Explain the difference between these two quantities.

Solution:

$$P(\text{Male} \mid \text{Salary} > \$125\text{K}) = \frac{15}{20} = 75\%,$$
$$P(\text{Salary} > \$125\text{K} \mid \text{Male}) = \frac{15}{26} \approx 58\%.$$

The quantity $P(\text{Male} \mid \text{Salary} > \$125\text{K})$ is the proportion of those listing salaries above $125K that are male. The quantity $P(\text{Salary} > \$125\text{K} \mid \text{Male})$ is the proportion of males listing salary above $125K.

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.

	Drop-off					
Pick-up	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 60 students enrolled in the class, 23 are female (38%) and 37 are male (62%). 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, $P(A) = \frac{37}{60}$, and $P(B \mid A) = \frac{36}{59}$. Thus, the probability that both will be male is

   $$\begin{aligned} P(A \cap B) &= P(A)P(B \mid A) \\ &= \frac{37}{60} \cdot \frac{36}{59} \\ &= \frac{1332}{3540} \\ &\approx 38\%. \end{aligned}$$

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

   Solution: Using the events A and B defined in the previous part, $P(A^c) = \frac{23}{60}$ and $P(B^c \mid A^c) = \frac{22}{59}$. Thus, the probability that both will be female is

   $$\begin{aligned} P(A^c \cap B^c) &= P(A^c)P(B^c \mid A^c) \\ &= \frac{23}{60} \cdot \frac{22}{59} \\ &= \frac{506}{3540} \\ &\approx 14\%. \end{aligned}$$

   (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,

$$\begin{aligned} P(A \cap B^c) &= P(A)P(B^c \mid A) \\ &= \frac{37}{60}\frac{23}{59} \\ &= \frac{851}{3540} \end{aligned}$$

$$\begin{aligned} P(A^c \cap B) &= P(A^c)P(B \mid A^c) \\ &= \frac{23}{60}\frac{37}{59} \\ &= \frac{851}{3540}. \end{aligned}$$

Thus,

$$\begin{aligned} P(\text{one male and one female}) &= \frac{851}{3540} + \frac{851}{3540} \\ &= \frac{1702}{3540} \\ &\approx 48\%. \end{aligned}$$

4. Of the 54 students who filled out the survey, 39 indicated that they drink at least one cup of coffee per day, while 15 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:
   $$\frac{39}{54} \cdot \frac{38}{53} = \frac{1482}{2862} \approx 52\%.$$

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

   Solution:
   $$\frac{15}{54} \cdot \frac{14}{53} = \frac{210}{2862} \approx 7\%.$$

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

   Solution:
   $$\frac{39}{54} \cdot \frac{15}{53} + \frac{15}{54} \cdot \frac{39}{53} = \frac{1170}{2862} \approx 41\%.$$

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 two dice. What is the probability of getting exactly one 6?

Solution: Define the following events:

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

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

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

These events are mutually exclusive, so

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

Using the independence of events A and B, we get

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

Note that these two expressions are equal. Thus,

$$\mathrm{P}(\text{exactly one } 6) = 2 \cdot \tfrac{1}{6} \cdot \tfrac{5}{6} \approx 28\%$$

7. Suppose that you sell fire insurance policies to two 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: 5%, and 1%. Assume that fire damages to the two buildings are independent events. Compute the probability that exactly one building gets damaged by fire in the next 10 years.

Solution:

$$(.05)(.99) + (.95)(.01) = .059 = 5.9\%$$

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

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

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.