

Conditional Probability

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

| Industry | Gender | | Total |
|---------------|--------|------|-------|
| | Female | Male | |
| 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.

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

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 | | | | | Total |
|------------|----------|----------|-----------|--------|------------|---------|
| | Bronx | Brooklyn | Manhattan | Queens | Staten Is. | |
| 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.

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

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?
 - (b) What is the probability that both students are female?
 - (c) What is the probability that one of the students is male and one of the students is female?

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?
 - (b) What is the probability that neither student regularly drinks coffee?
 - (c) What is the probability that exactly one student regularly drinks coffee?

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 $P(A \cap B)$.
6. Suppose that you roll three dice. What is the probability of getting exactly one 6?
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.
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 = the first review is 4 or 5 stars

B = 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.
- (b) You randomly sample one business from the dataset, then sample two distinct reviews of the business.
- (c) You randomly sample one user from the dataset, then sample two distinct reviews written by the user.

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.
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?
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?