Motivating Probability Through Contingency Tables

Reading: Chapter 3.1-3.2

Important concepts:

- A **random process** is a process where we know the set of all possible **outcomes** (or **events**), but we do not know the order in which they occur.
- The **probability** of outcome is the proportion of times the outcome would occur if we observed the random process an infinite number of times.
- Two outcomes are **disjoint** (also known as mutually exclusive) if they cannot both happen.
- A **probability distribution** is a table of all disjoint outcomes and their associated probabilities.
  - The outcomes listed are disjoint.
  - Each probability (of an outcome) must be between 0 and 1.
  - The probabilities (of disjoint events) must total 1.
- The **complement** of an event represents all outcomes not in that event.
- Two processes are **independent** if knowing the outcome of one provides no useful information about the outcome of the other.
- A probability based on a single event is a **marginal probability**.
- A probability based on two or more events is a **joint probability**.
- A probability based on two events, where one is the outcome of interest, and the other is the condition, is a **conditional probability**.

Probability rules:

- Addition rule of *disjoint* outcomes: \$P(A_1 or A_2)=P(A_1)+P(A_2) \$ if events A_1 and A_2 are disjoint
- General addition rule: $P(A_1 \text{ or } A_2) = P(A_1) + P(A_2) - P(A_1 \text{ and } A_2)$
- Complement rule: $P(A) = 1 - P(A^c)$
- Multiplication rule for *independent* processes: If events A_1 and A_2 are independent, $P(A_1 \text{ and } A_2) = P(A_1) \times P(A_2)$.
- Conditional probability: $P(A_1|A_2) = \frac{P(A_1 \text{ and } A_2)}{P(A_2)}$
- General multiplication rule: For any events A_1 and A_2, $P(A_1 \text{ and } A_2) = P(A_1|A_2)P(A_2)$
- Sum of conditional probabilities: $P(A_1|A_2) = 1 - P(A_1^c|A_2)$

Example 1: Firstie Plants

One hundred Mount Holyoke students sent their firstie plants home with (one hundred different) friends for the summer. Regular watering over the course of the summer is assumed to be important for plant survival. At the end of the summer, information about the plants' watering and survival were recorded in the following table:

	Alive	Dead	Total
Watered Regularly	70	5	75
Not Watered Regularly	10	15	25
Total	80	20	100

Suppose that these data are representative of what could happen to your firstie plant if you send it home with a friend for the summer. For the following questions, assume that you send your plant home with a friend over the summer.

1) What is the probability that your plant will be alive at the end of the summer?

- Math:
- Type of probability:
- Notation:
- Complete answer:

2) What is the probability that your plant is alive at the end of the summer and was regularly watered?

- Math:
- Type of probability:
- Notation:
- Complete answer:

3) If your plant is dead at the end of the summer, what's the chance that your friend didn't water it regularly?

- Math:
- Type of probability:
- Notation:
- Complete answer:

4) If your friend didn't water your plant regularly, what's the chance it'll be dead at the end of the summer?

- Math:
- Type of probability:
- Notation:
- Complete answer:

5) What is the difference between (3) and (4)?

6) Give an example of disjoint events.

7) Is survival status dependent on whether a plant is regularly watered?

8) What probabilities can be found in a contingency table?

- A: alive
- A^c: not alive
- W: regularly watered
- W^c: not regularly watered

	Alive	Dead	Total
Watered Regularly	70/100	5/100	75/100
Not Watered Regularly	10/100	15/100	25/100
Total	80/100	20/100	100/100

	Alive	Dead	Total
Watered Regularly	$P($ $)$	$P($ $)$	$P($ $)$
Not Watered Regularly	$P($ $)$	$P($ $)$	$P($ $)$
Total	$P($ $)$	$P($ $)$	1

Example 2: Breast cancer

We want to estimate the probability that a woman with a positive mammogram actually has breast cancer, even though she's in a low-risk group: 40 to 50 years old, with no symptoms or family history of breast cancer. The probability that one of these women has breast cancer is 0.8 percent. If a woman has breast cancer, the probability is 90 percent that she will have a positive mammogram. If a woman does not have breast cancer, the probability is 7 percent that she will still have a positive mammogram. Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

Discussion: https://opinionator.blogs.nytimes.com/2010/04/25/chances-are/

Approach 1: Translate these words into counts. For purposes of this translation, let's suppose these numbers are based on 1000 people.

Workspace:

Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?

Approach 2: Translate these words into their corresponding probabilities:

(a) What is the probability that a woman with a positive mammogram actually has breast cancer, even though she's in a low-risk group: 40 to 50 years old, with no symptoms or family history of breast cancer?

- Type of probability:
- Notation:
- Complete answer:

(b) Given that a woman (in this group) has breast cancer, what is the probability that they have a positive mammogram?

- Type of probability:
- Notation:
- Complete answer:

(c) Given that a woman (in this group) does not have breast cancer, what is the probability that they have a positive mammogram?

- Type of probability:
- Notation:
- Complete answer:

(d) Imagine a woman who has a positive mammogram. What is the probability that she actually has breast cancer?