8 Probability Rules

8.1 Multiplication rule

- Rearranging the definition of conditional probability we get the **Multiplication rule:** the probability that two events both occur is

$$P(A \cap B) = P(A|B)P(B)$$
$$= P(B|A)P(A)$$

- The multiplication rule says that you should think "multiply" when you see "and".
- However, be careful about *what* you are multiplying: to find a joint probability you need an unconditional and an appropriate conditional probability.
- You can condition either on A or on B, provided you have the appropriate marginal probability; often, conditioning one way is easier than the other.
- Be careful: the multiplication rule does *not* say that $P(A \cap B)$ is the same as $P(A)P(B)$.

- The multiplication rule is useful in situations where conditional probabilities are easier to obtain directly than joint probabilities.

Example 8.1. A standard deck of playing cards has 52 cards, 13 cards (2 through 10, jack, king, queen, ace) in each of 4 suits (hearts, diamonds, clubs, spades). Shuffle a deck and deals cards one at a time without replacement.

1. Find the probability that the first card dealt is a heart.

2. If the first card dealt is a heart, determine the conditional probability that the second card is a heart.

3. Find the probability that the first two cards dealt are hearts.

4. Find the probability that the first two cards dealt are hearts and the third card dealt is a diamond.

- The multiplication rule extends naturally to more than two events (though the notation gets messy). For three events, we have

$$P(A_1 \cap A_2 \cap A_3) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)$$

- And in general,

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap \cdots) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_4) \cdots$$

- The multiplication rule is useful for computing probabilities of events that can be broken down into component "stages" where conditional probabilities at each stage are readily available. At each stage, condition on the information about all previous stages.

Example 8.2. The birthday problem concerns the probability that at least two people in a group of n people have the same birthday[1]. Ignore multiple births and February 29 and assume that the other 365 days are all equally likely[2].

1. If $n = 30$, what do you think the probability that at least two people share a birthday is: 0-20%, 20-40%, 40-60%, 60-80%, 80-100%? How large do you think n needs to be in order for the probability that at least two people share a birthday to be larger than 0.5? Just make guesses before proceeding to calculations.

---

[1]You should really click on this birthday problem link.

[2]Which isn't quite true. However, a non-uniform distribution of birthdays only increases the probability that at least two people have the same birthday. To see that, think of an extreme case like if everyone were born in September.

2. Now consider $n = 3$ people, labeled 1, 2, and 3. What is the probability that persons 1 and 2 have different birthdays?

3. What is the probability that persons 1, 2, and 3 all have different birthdays *given* that persons 1 and 2 have different birthdays?

4. What is the probability that persons 1, 2, and 3 all have different birthdays?

5. When $n = 3$. What is the probability that at least two people share a birthday?

6. For $n = 30$, find the probability that none of the people have the same birthday.

7. For $n = 30$, find the probability that at least two people have the same birthday.

8. Write a clearly worded sentence interpreting the probability in the previous part as a long run relative frequency.

9. When $n = 100$ the probability is about 0.9999997.

   If you are in a group of 100 people and no one shares your birthday, should you be surprised? Discuss.

8.2 Law of total probability

Example 8.3. Suppose that

- 67% of Democrats believe in human-driven climate
- 46% of Independents believe in human-driven climate
- 34% of Republicans believe in human-driven climate

Also suppose that

- 28% of American adults are Democrats
- 42% of American adults are Independents
- 30% of American adults are Republicans

Randomly select an American adult. Compute the probability that the selected person believes in human-driven climate change.

- **Law of total probability.** If $C_1, \dots, C_k$ are disjoint with $C_1 \cup \cdots \cup C_k = \Omega$, then

$$\mathrm{P}(A) = \sum_{i=1}^{k} \mathrm{P}(A \cap C_i)$$
$$= \sum_{i=1}^{k} \mathrm{P}(A|C_i)\mathrm{P}(C_i)$$

- The events $C_1, \dots, C_k$, which represent the "cases", form a *partition* of the sample space; each outcome $\omega \in \Omega$ lies in exactly one of the C_i.
- The law of total probability says that we can interpret the unconditional probability $\mathrm{P}(A)$ as a probability-weighted average of the case-by-case conditional probabilities $\mathrm{P}(A|C_i)$ where the weights $\mathrm{P}(C_i)$ represent the probability of encountering each case.

- Conditioning and using the law of probability is an effective strategy in solving many problems, even when the problem doesn't seem to involve conditioning.
- For example, when a problem involves iterations or steps it is often useful to *condition on the result of the first step*.

Example 8.4. You and your friend are playing the "lookaway challenge".

The game consists of possibly multiple rounds. In the first round, you point in one of four directions: up, down, left or right. At the exact same time, your friend also looks in one of those four directions. If your friend looks in the same direction you're pointing, you win! Otherwise, you switch roles and the game continues to the next round — now your friend points in a direction and you try to look away. As long as no one wins, you keep switching off who points and who looks. The game ends, and the current "pointer" wins, whenever the "looker" looks in the same direction as the pointer.

Suppose that each player is equally likely to point/look in each of the four directions, independently from round to round. What is the probability that you win the game?

1. Why might you expect the probability to not be equal to 0.5?

2. If you start as the pointer, what is the probability that you win in the first round?

3. If p denotes the probability that the player who starts as the pointer wins the game, what is the probability that the player who starts as the looker wins the game? (Note: p is the probability that the person who starts as pointer wins the whole game, not just the first round.)

4. Let A be the event that the person who starts as the pointer wins the game, and B be the event that the person who starts as the pointer wins in the first round. What is $P(A|B)$?

5. Find a simple expression for $P(A|B^c)$ in terms of p. The key is to consider this question: if the player who starts as the pointer does not win in the first round, how does the game behave from that point forward?

6. Condition on the result of the first round and set up an equation to solve for p.

7. Interpret the probability from the previous part.

8.3 Bayes Rule

Example 8.5. Suppose that

- 67% of Democrats believe in human-driven climate
- 46% of Independents believe in human-driven climate
- 34% of Republicans believe in human-driven climate

Also suppose that

- 28% of American adults are Democrats
- 42% of American adults are Independents
- 30% of American adults are Republicans

Randomly select an American adult. Compute the conditional probability that the selected adult is a Democrat given that they believe in human-driven climate change.

- **Bayes' rule for events** specifies how a prior probability $P(H)$ of event H is updated in response to the evidence E to obtain the posterior probability $P(H|E)$.

$$\text{P}(H|E) = \frac{\text{P}(E|H)\text{P}(H)}{\text{P}(E)}$$

- Event H represents a particular hypothesis (or model or case)
- Event E represents observed evidence (or data or information)
- $\text{P}(H)$ is the unconditional or **prior probability** of H (prior to observing evidence E)
- $\text{P}(H|E)$ is the conditional or **posterior probability** of H after observing evidence E.
- $\text{P}(E|H)$ is the **likelihood** of evidence E given hypothesis (or model or case) H

Example 8.6. Continuing the previous example. Randomly select an American adult.

1. Consider the conditional probability that the selected adult is a Democrat given that they believe in human-driven climate change. Identify the prior probability, hypothesis, evidence, likelihood, and posterior probability.

2. Compute the conditional probability that the selected adult is a Republican given that they believe in human-driven climate change.

3. How many times more likely is it for an American adult to be a Democrat than be be a Republican?

4. How many times more likely is it for an American adult to believe in human-driven climate change given that they are a Democrat than given that they are a Republican?

5. How many times more likely is it for an American adult who believes in human-driven climate change to be a Democrat than to be a Republican?

6. How are the values in the three previous parts related?

7. Compute the conditional probability that the selected adult is an Independent given that they believe in human-driven climate change.

8. How many times more likely is it for an American adult to be an Independent than be be a Republican?

9. How many times more likely is it for an American adult to believe in human-driven climate change given that they are an Independent than given that they are a Republican?

10. How many times more likely is it for an American adult who believes in human-driven climate change to be an Independent than to be a Republican?

11. How are the values in the three previous parts related?

- Bayes rule is often used when there are multiple hypotheses or cases. Suppose $H_1, \ldots, H_k$ is a series of distinct hypotheses which together account for all possibilities, and E is any event (evidence).
- Combining Bayes' rule with the law of total probability,

$$
\begin{aligned}
P(H_j|E) &= \frac{P(E|H_j)P(H_j)}{P(E)} \\
&= \frac{P(E|H_j)P(H_j)}{\sum_{i=1}^{k} P(E|H_i)P(H_i)}
\end{aligned}
$$

$$
P(H_j|E) \propto P(E|H_j)P(H_j)
$$

- The symbol $\propto$ is read "is proportional to". The relative *ratios* of the posterior probabilities of different hypotheses are determined by the product of the prior probabilities and the likelihoods, $P(E|H_j)P(H_j)$. The marginal probability of the evidence, $P(E)$, in the denominator simply normalizes the numerators to ensure that the updated conditional probabilities given the evidence sum to 1 over all the distinct hypotheses.
- **In short, Bayes' rule says**

$$\textbf{posterior} \propto \textbf{likelihood} \times \textbf{prior}$$

Example 8.7. Suppose that you are presented with six boxes, labeled 0, 1, 2, ..., 5, each containing five marbles. Box 0 contains 0 green and 5 gold marbles, box 1 contains 1 green and 4 gold, and so on with box i containing i green and $5 - i$ gold. One of the boxes is chosen uniformly at random (perhaps by rolling a fair six-sided die), and then you will randomly select marbles from that box, without replacement. Imagine the boxes appear identical and you can't see inside; all you observe is the color of the marbles you select. Based on the colors of the marbles selected, you will update the probabilities of which box had been chosen.

1. Suppose that a single marble is selected and it is green. Which box do you think is the most likely to have been chosen? Make a guess for the posterior probabilities for each box. Then construct a Bayes table to compute the posterior probabilities. How do they compare to the prior probabilities?

2. Now suppose a second marble is selected from the same box, without replacement, and its color is gold. Which box do you think is the most likely to have been chosen given these two marbles? Make a guess for the posterior probabilities for each box. Then construct a Bayes table to compute the posterior probabilities, *using the posterior probabilities from the previous part after the selection of the green marble as the new prior probabilities before seeing the gold marble.*

3. Now construct a Bayes table corresponding to the original prior probabilities (1/6 each) and the combined evidence that the first ball selected was green and the second was gold. How do the posterior probabilities compare to the previous part?

- Bayesian analysis is often an iterative process.
- Posterior probabilities are updated after observing some information or data. These probabilities can then be used as prior probabilities before observing new data.
- Posterior probabilities can be sequentially updated as new data becomes available, with the posterior probabilities after the previous stage serving as the prior probabilities for the next stage.
- The final posterior probabilities only depend upon the cumulative data. It doesn't matter if we sequentially update the posterior after each new piece of data or only once after all the data is available; the final posterior probabilities will be the same either way. Also, the final posterior probabilities are not impacted by the order in which the data are observed.