Conditional Probability

Conditional probability gives us a systematic way to reason about the outcome of an experiment based on *partial information*.

Examples:

- A die is rolled two times. You are told that the sum of the rolls is 9. What is the probability that the first roll was a 6?
- The alarm is going off in my house. What is the probability that there is a burglar present?
- You enjoyed watching "The Avengers". How likely is it that you will enjoy watching "Iron Man 3"?

Given knowledge of an event B, we can construct a new (updated) probability law for outcomes in Ω.

Definition: The **conditional probability** of an event A given the occurrence of an event B with $\mathrm{P}(B) > 0$ is

$$\mathrm{P}(A|B) \;=\; \frac{\mathrm{P}(A \cap B)}{\mathrm{P}(B)}.$$

It is not hard to check that if $\mathrm{P}(\cdot)$ is a valid probability law on Ω (i.e., satisfies the Kolmogorov axioms), then $\mathrm{P}(\cdot|B)$ is also a valid probability law.

Important Note: $\mathrm{P}(A|B) \neq \mathrm{P}(B|A)$. This is one of the most common mistakes made by people with no background in probability. Much more on this later.

Example. Recall the example where we toss a fair coin three consecutive times. The sample space consists of eight sequences:

$$\Omega = \{\text{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT}\}.$$

Now suppose that we wish to find the conditional probability $\mathrm{P}(A|B)$ where A and B are the events

$$\begin{aligned} A &= \{\text{we toss two or more heads}\}, \\ B &= \{\text{the first toss is a head}\}. \end{aligned}$$

First compute

$$\mathrm{P}(B) =$$

Then find

$$\mathrm{P}(A \cap B) =$$

And then combine these to find

$$\mathrm{P}(A|B) =$$

Compare this to $\mathrm{P}(A)$.

Exercise:
Suppose that we now toss a fair coin ten consecutive times. Let A and B denote the events

$$A = \{\text{we toss two or more heads}\},$$
$$B = \{\text{the first toss is a head}\}.$$

What is the conditional probability $\mathrm{P}(A|B)$?

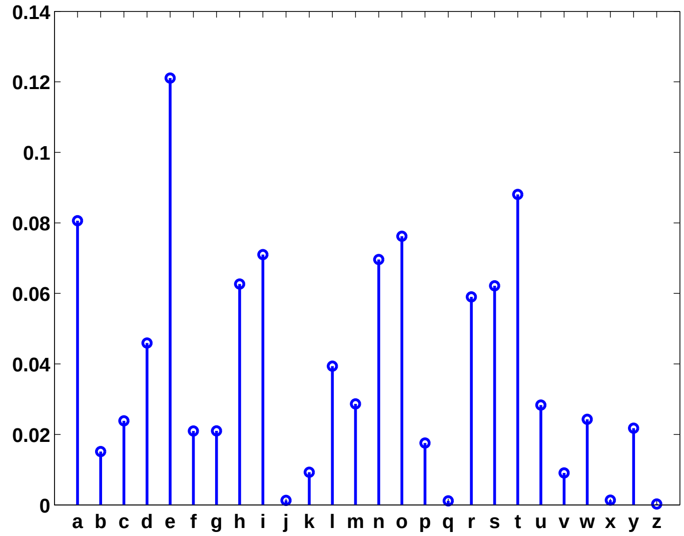
Exercise:
Let (X, Y) be a point in the unit square $[0, 1] \times [0, 1]$ generated using the uniform probability law. Let events A and B be defined as

$$A = \{X + Y \leq 1\}, \qquad B = \{Y \leq 1/2\}.$$

(a) Sketch the sample space Ω and the events A and B as subsets of Ω.

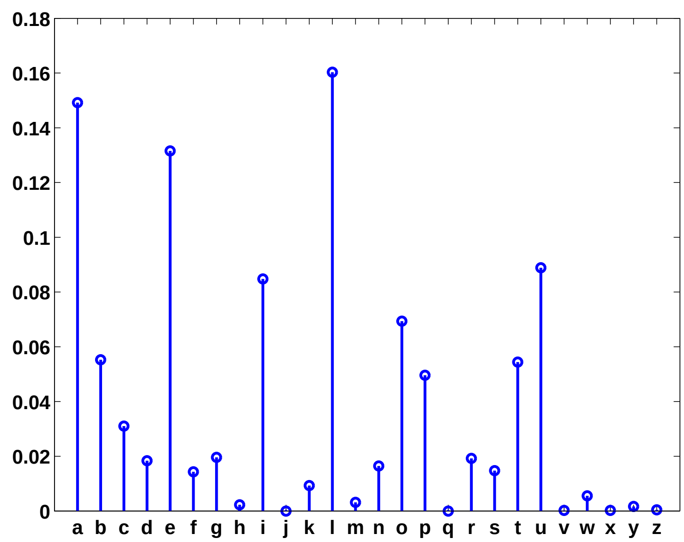
(b) Calculate $\mathrm{P}(A|B)$.

Example. Suppose I open a book written in English, put my finger down at a random location, and then note which letter is closest. Here are the probabilities for each of the 26 letters:



(Another way to interpret this: A's make up just over 8% of letters in English, B's make up about 1.8%, E's about 12.5%, etc.)

Now suppose that I see my finger falls on an L. Here are the conditional probabilities for the letter that immediately following the L in the text:



Independence and Conditional Probability

Conditional probability gives us a very nice way to interpret (and check for) the independence of two events. Events A and B are **independent** if and only if

$$\mathrm{P}(A|B) = \mathrm{P}(A) \quad \text{and/or} \quad \mathrm{P}(B|A) = \mathrm{P}(B).$$

(We say "and/or" above since if one of those relations is true, the other must be, so we only need to check for one of them.)

This is a mathematical formalization of the idea that if A and B are independent, then learning about A tells us nothing about B (and vice versa) — the probability that it occurs does not change.

Actually, the exercises on pages 23–25 are much easier using this definition rather than checking $\mathrm{P}(A \cap B) = \mathrm{P}(A)\,\mathrm{P}(B)$ directly.

The Multiplication Rule

Conditional probability can help us break down complicated probability calculations into a series of manageable pieces. Suppose we are interested in the probability of a sequence of events $A_1, A_2, \ldots, A_n$ all occurring. That is, we wish to compute

$$\mathrm{P}\left(A_1 \cap A_2 \cap A_3 \cap \cdots \cap A_n\right) = \mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right).$$

We can break down this calculation using conditional probability as follows:

$$\begin{aligned}
\mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right) &= \mathrm{P}\left(A_1\right) \cdot \frac{\mathrm{P}\left(A_1 \cap A_2\right)}{\mathrm{P}\left(A_1\right)} \cdot \frac{\mathrm{P}\left(A_1 \cap A_2 \cap A_3\right)}{\mathrm{P}\left(A_1 \cap A_2\right)} \cdots \frac{\mathrm{P}\left(\bigcap_{i=1}^{n} A_i\right)}{\mathrm{P}\left(\bigcap_{j=1}^{n-1} A_j\right)} \\
&= \mathrm{P}\left(A_1\right)\, \mathrm{P}\left(A_2 | A_1\right)\, \mathrm{P}\left(A_3 | A_2 \cap A_1\right) \cdots \\
&\qquad \mathrm{P}\left(A_n | A_1 \cap A_2 \cap \ldots \cap A_{n-1}\right)
\end{aligned}$$

A few examples will make it clear why this expansion is useful.

Example. We draw 3 cards from an ordinary 52 card deck. What is the probability that none of the three is a heart?

To solve this, we let

$$\begin{aligned} A_1 &= \{\text{the 1st card is not a heart}\} \\ A_2 &= \{\text{the 2nd card is not a heart}\} \\ A_3 &= \{\text{the 3rd card is not a heart}\} \end{aligned}$$

We want to compute $\mathrm{P}(A_1 \cap A_2 \cap A_3)$. Applying the multiplication rule,

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\ \mathrm{P}(A_2|A_1)\ \mathrm{P}(A_3|A_1 \cap A_2).$$

Computing each of the three terms on the right is straightforward. Of the 52 cards in the deck, 39 are not hearts and so

$$\mathrm{P}(A_1) = \frac{39}{52} \quad \left(= \frac{3}{4}\right).$$

Given that the first card was not a heart, we know that 38 of the remaining 51 cards are not hearts, so

$$\mathrm{P}(A_2|A_1) = \frac{38}{51}.$$

Given that the first two cards were not hearts, we know that 37 of the remaining 50 cards are not hearts, and so

$$\mathrm{P}(A_3|A_1 \cap A_2) = \frac{37}{50}.$$

Thus

$$\mathrm{P}(A_1 \cap A_2 \cap A_3) = \frac{39 \cdot 38 \cdot 37}{52 \cdot 51 \cdot 50} \approx 0.4135.$$

Exercise:
We have 16 marbles, 12 of which are white, and 4 of which are black. We randomly divide them into 4 groups of 4 marbles. What is the probability that each group contains exactly one black marble?

You can attack this problem with the multiplication rule by labeling the black marbles with numbers $1, 2, 3, 4$, and then defining the events

$$A_1 = \{\text{marbles 1 and 2 are in different groups}\}$$
$$A_2 = \{\text{marbles 1, 2, and 3 are in different groups}\}$$
$$A_3 = \{\text{marbles 1, 2, 3, and 4 are in different groups}\}.$$

Of course, we are interested in $\mathrm{P}(A_3)$, but since $A_3 \subset A_2 \subset A_1$, we have

$$A_3 = A_1 \cap A_2 \cap A_3,$$

and so

$$\mathrm{P}(A_3) = \mathrm{P}(A_1 \cap A_2 \cap A_3) = \mathrm{P}(A_1)\,\mathrm{P}(A_2|A_1)\,\mathrm{P}(A_3|A_1 \cap A_2).$$