4.2: Conditional Probability

We have described the whole foundation of the theory of probability as coming from *imperfect knowledge*, in the sense that we don't know for sure if an event A will happen any particular time we do the experiment but we do know, in the long run, in what fraction of times A will happen. Or, at least, we claim that there is some number $P(A)$ such that after running the experiment N times, out of which n_A of these times are when A happened, $P(A)$ is approximately n_A/N (and this ratio gets closer and closer to $P(A)$ as N gets bigger and bigger).

But what if we have *some* knowledge? In particular, what happens if we know for sure that the event B has happened – will that influence our knowledge of whether A happens or not? As before, when there is randomness involved, we cannot tell for sure if A will happen, but we hope that, given the knowledge that B happened, we can make a more accurate guess about the probability of A.

[eg:condprob1] If you pick a person at random in a certain country on a particular date, you might be able to estimate the probability that the person had a certain height if you knew enough about the range of heights of the whole population of that country. [In fact, below we will make estimates of this kind.] That is, if we define the event

$$A = \text{``the random person is taller than } 1.829 \text{ meters (6 feet)''} \tag{4.2.1}$$

then we might estimate $P(A)$.

But consider the event

$$B = \text{``the random person's parents were both taller than } 1.829 \text{ meters''} \ . \tag{4.2.2}$$

Because there is a genetic component to height, if you know that B happened, it would change your idea of how likely, given that knowledge, that A happened. Because genetics are not the only thing which determines a person's height, you would not be certain that A happened, even given the knowledge of B.

Let us use the frequentist approach to derive a formula for this kind of *probability of A given that B is known to have happened*. So think about doing the repeatable experiment many times, say N times. Out of all those times, some times B happens, say it happens n_B times. Out of *those* times, the ones where B happened, sometimes A also happened. These are the cases where both A and B happened – or, converting this to a more mathematical descriptions, the times that $A \cap B$ happened – so we will write it $n_{A \cap B}$.

We know that the probability of A happening in the cases where we know for sure that B happened is approximately $n_{A\cap B}/n_B$. Let's do that favorite trick of multiplying and dividing by the same number, so finding that the probability in which we are interested is approximately

$$\frac{n_{A\cap B}}{n_B} = \frac{n_{A\cap B}\cdot N}{N\cdot n_B} = \frac{n_{A\cap B}}{N}\cdot\frac{N}{n_B} = \frac{n_{A\cap B}}{N}\Bigg/\frac{n_B}{N} \approx P(A\cap B)\Big/P(B) \tag{4.2.3}$$

Which is why we make the

[def:condprob] The **conditional probability** is

$$P(A|B) = \frac{P(A\cap B)}{P(B)} \ . \tag{4.2.4}$$

Here $P(A|B)$ is pronounced *the probability of A given B*.

Let's do a simple

EXAMPLE 4.2.3. Building off of Example 4.1.19, note that the probability of rolling a 2 is $P(\{2\}) = 1/6$ (as is the probability of rolling any other face – it's a *fair die*). But suppose that you were told that the roll was even, which is the event $\{2, 4, 6\}$, and asked for the probability that the roll was a 2 given this prior knowledge. The answer would be

$$P(\{2\} \mid \{2,4,6\}) = \frac{P(\{2\}\cap\{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\{2\})}{P(\{2,4,6\})} = \frac{1/6}{1/2} = 1/3 \ . \tag{4.2.5}$$

In other words, the probability of rolling a 2 on a fair die with no other information is $1/6$, which the probability of rolling a 2 given that we rolled an even number is $1/3$. So the probability doubled with the given information.

Sometimes the probability changes even more than merely doubling: the probability that we rolled a 1 with no other knowledge is $1/6$, while the probability that we rolled a 1 given that we rolled an even number is

$$P(\{1\} \mid \{2,4,6\}) = \frac{P(\{1\} \cap \{2,4,6\})}{P(\{2,4,6\})} = \frac{P(\emptyset)}{P(\{2,4,6\})} = \frac{0}{1/2} = 0\,. \tag{4.2.6}$$

But, actually, sometimes the conditional probability for some event is the same as the unconditioned probability. In other words, sometimes knowing that B happened doesn't change our estimate of the probability of A at all, they are no really related events, at least from the point of view of probability. This motivates the

[def:independent] Two events A and B are called **independent** if $P(A \mid B) = P(A)$.

Plugging the defining formula for $P(A \mid B)$ into the definition of *independent*, it is easy to see that

FACT 4.2.5. Events A and B are independent if and only if $P(A \cap B) = P(A) \cdot P(B)$.

EXAMPLE 4.2.6. Still using the situation of Example 4.1.19, we saw in Example 4.2.3 that the events $\{2\}$ and $\{2,3,4\}$ are not independent since

$$P(\{2\}) = 1/6 \neq 1/3 = P(\{2\} \mid \{2,4,6\}) \tag{4.2.7}$$

nor are $\{1\}$ and $\{2,3,4\}$, since

$$P(\{1\}) = 1/6 \neq 0 = P(\{1\} \mid \{2,4,6\})\,. \tag{4.2.8}$$

However, look at the events $\{1,2\}$ and $\{2,4,6\}$:

$$\begin{aligned} P(\{1,2\}) = P(\{1\}) + P(\{2\}) &= 1/6 + 1/6 \\ &= 1/3 \\ &= \frac{1/6}{1/2} \\ &= \frac{P(\{1\})}{P(\{2,4,6\})} \\ &= \frac{P(\{1,2\} \cap \{2,4,6\})}{P(\{2,4,6\})} \\ &= P(\{1,2\} \mid \{2,4,6\}) \end{aligned}$$

which means that they are independent!

EXAMPLE 4.2.7. We can now fully explain what was going on in Example 4.1.21. The two fair dice were supposed to be rolled in a way that the first roll had no effect on the second – this exactly means that the dice were rolled *independently*. As we saw, this then means that each individual outcome of sample space S had probability $\frac{1}{36}$. But the first roll having any particular value is independent of the second roll having another, *e.g.*, if $A = \{11,12,13,14,15,16\}$is the event in that sample space of getting a 1 on the first roll and $B = \{14,24,34,44,54,64\}$is the event of getting a 4 on the second roll, then events A and B are independent, as we check by using Fact 4.2.5:

$$\begin{aligned} P(A \cap B) &= P(\{14\}) \\ &= \frac{1}{36} \\ &= \frac{1}{6} \cdot \frac{1}{6} \\ &= \frac{6}{36} \cdot \frac{6}{36} \\ &= P(A) \cdot P(B)\,. \end{aligned}$$

On the other hand, the event "the sum of the rolls is 4," which is $C = \{13,22,31\}$as a set, *is not independent* of the value of the first roll, since $P(A \cap C) = P(\{13\}) = \frac{1}{36}$ but $P(A) \cdot P(C) = \frac{6}{36} \cdot \frac{3}{36} = \frac{1}{6} \cdot \frac{1}{12} = \frac{1}{72}$.