2.2 Definition and intuition

Unit 2: Conditioning

Adapted from Blitzstein-Hwang Chapter 2.

Definition 2.2.1 (Conditional probability).

If A and B are events with $P(B) > 0$, then the *conditional probability* of A given B, denoted by $P(A|B)$, is defined as

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

Here A is the event whose uncertainty we want to update, and B is the evidence we observe (or want to treat as given). We call $P(A)$ the *prior* probability of A and $P(A|B)$ the *posterior* probability of A ("prior" means before updating based on the evidence, and "posterior" means after updating based on the evidence). It is important to interpret the event appearing after the vertical conditioning bar as the evidence that we have observed or that is being conditioned on: $P(A|B)$ is the probability of A given the evidence B, *not* the probability of some entity called $A|B$.

Example 2.2.2 (Two cards).

A standard deck of 52 playing cards is shuffled well. Two cards are drawn randomly, one at a time without replacement. Let A be the event that the first card is a heart, and B be the event that the second card is red. Find $P(A|B)$ and $P(B|A)$.

Solution

By the naive definition of probability and the multiplication rule,

$$P(A \cap B) = \frac{13 \cdot 25}{52 \cdot 51} = \frac{25}{204},$$

since a favorable outcome is determined by choosing any of the 13 hearts and then any of the remaining 25 red cards. Also, $P(A) = 1/4$ since the 4 suits are equally likely, and

$$P(B) = \frac{26 \cdot 51}{52 \cdot 51} = \frac{1}{2}$$

since there are 26 favorable possibilities for the *second* card, and for each of those, the first card can be any other card (recall that chronological order is not needed in the multiplication rule). A neater way to see that $P(B) = 1/2$ is by *symmetry*: from a vantage point before having done the experiment, the second card is equally likely to be any card in the deck. We now have all the pieces needed to apply the definition of conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{25/204}{1/2} = \frac{25}{102},$$
$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{25/204}{1/4} = \frac{25}{51}.$$

This is a simple example, but already there are several things worth noting.

1. It's extremely important to be careful about which events to put on which side of the conditioning bar. In particular, $P(A|B) \neq P(B|A)$. The next section explores how $P(A|B)$ and $P(B|A)$ are related in general. Confusing these two quantities is called the *prosecutor's fallacy*. If instead we had defined B to be the event that the second card is a heart, the two conditional probabilities would have been equal.

2. Both $P(A|B)$ and $P(B|A)$ make sense (intuitively and mathematically); the chronological order in which cards were chosen does not dictate which conditional probabilities we can look at. When we calculate conditional probabilities, we are considering what *information* observing one event provides about another event, not whether one event *causes* another.

To shed more light on what conditional probability means, here are two intuitive interpretations.

Intuition 2.2.3 (Pebble World).

Consider a finite sample space, with the outcomes visualized as pebbles with total mass 1. Since A is an event, it is a set of pebbles, and likewise for B. Figure 2.2.4(a) shows an example.

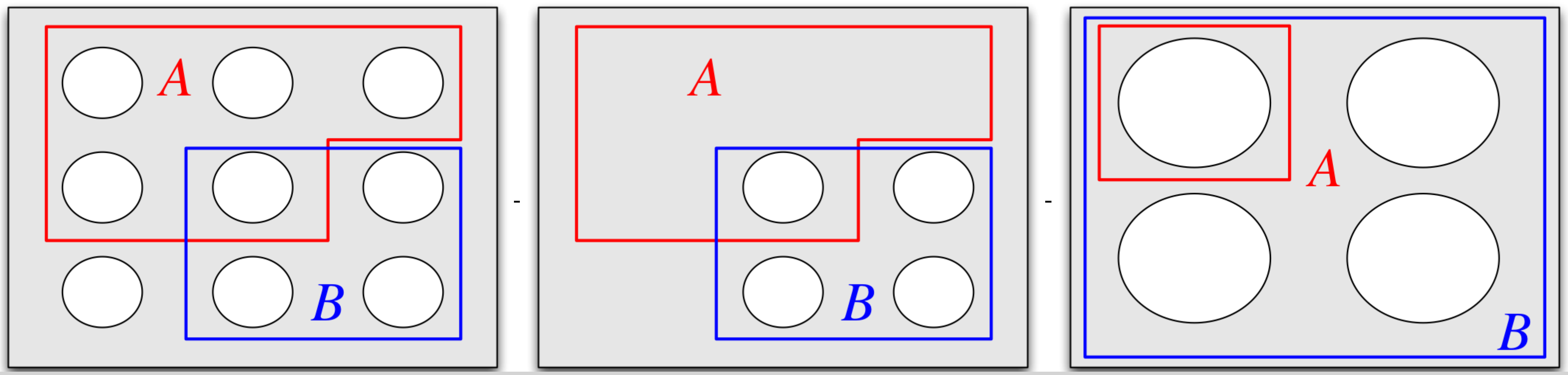

Figure 2.2.4: Pebble World intuition for $P(A|B)$. From left to right:
(a) Events A and B are subsets of the sample space. View Larger Image. Image Description.
(b) Because we know B occurred, get rid of the outcomes in B^c. View Larger Image. Image Description.
(c) In the restricted sample space, renormalize so the total mass is still 1. View Larger Image. Image Description.

Now suppose that we learn that B occurred. In Figure 2.2.4(b), upon obtaining this information, we get rid of all the pebbles in B^c because they are incompatible with the knowledge that B has occurred. Then $P(A \cap B)$ is the total mass of the pebbles remaining in A. Finally, in Figure 2.2.4(c), we *renormalize*, that is, divide all the masses by a constant so that the new total mass of the remaining pebbles is 1. This is achieved by dividing by $P(B)$, the total mass of the pebbles in B. The updated mass of the outcomes corresponding to event A is the conditional probability $P(A|B) = P(A \cap B)/P(B)$.

Intuition 2.2.5 (Frequentist interpretation).

Imagine repeating an experiment many times, randomly generating a long list of observed outcomes, each of them represented by a string of twenty-four 0's and 1's. The conditional probability of A given B can then be thought of in a natural way: it is the fraction of times that A occurs, restricting attention to the trials where B occurs. In Figure 2.2.6, our experiment has outcomes which can be represented as a string of 0's and 1's; B is the event that the first digit is 1 and A is the event that the second digit is 1. Conditioning on B, we circle all the repetitions where B occurred, and then we look at the fraction of circled repetitions in which event A also occurred.

In symbols, let n_A, n_B, n_{AB} be the number of occurrences of $A, B, A \cap B$ respectively in a large number n of repetitions of the experiment. The frequentist interpretation is that

$$P(A) \approx \frac{n_A}{n}, P(B) \approx \frac{n_B}{n}, P(A \cap B) \approx \frac{n_{AB}}{n}.$$

Then $P(A|B)$ is interpreted as n_{AB}/n_B, which equals $(n_{AB}/n)/(n_B/n)$. This interpretation again translates to $P(A|B) = P(A \cap B)/P(B)$.

10101010101001000010010

01010101001010000111110

11110010101010000000111

11001101101010101010110

01100001110101011110011

10100110111001001001110

•
•
•

Figure 2.2.6: Frequentist intuition for $P(A|B)$. The repetitions where B occurred are circled; among these, the repetitions where A occurred are highlighted in bold. $P(A|B)$ is the long-run relative frequency of the repetitions where A occurs, within the subset of repetitions where B occurs.

View Larger Image

Image Description