2.3 Bayes' rule and the law of total probability

Unit 2: Conditioning

Adapted from Blitzstein-Hwang Chapter 2.

The definition of conditional probability is simple—just a ratio of two probabilities—but it has far-reaching consequences. The first consequence is obtained easily by moving the denominator in the definition to the other side of the equation.

> THEOREM 2.3.1.
>
> For any events A and B with positive probabilities,
>
> $$P(A \cap B) = P(B)P(A|B) = P(A)P(B|A).$$

This follows from taking the definition of $P(A|B)$ and multiplying both sides by $P(B)$, and then taking the definition of $P(B|A)$ and multiplying both sides by $P(A)$. At first sight this theorem may not seem very useful: it *is* the definition of conditional probability, just written slightly differently, and anyway it seems circular to use $P(A|B)$ to help find $P(A \cap B)$ when $P(A|B)$ was defined in terms of $P(A \cap B)$. But we will see that the theorem is in fact very useful, since it often turns out to be possible to find conditional probabilities without going back to the definition, and in such cases Theorem 2.3.2 can help us more easily find $P(A \cap B)$.

Applying Theorem 2.3.1 repeatedly, we can generalize to the intersection of n events.

> THEOREM 2.3.2.
>
> For any events $A_1, \ldots, A_n$ with positive probabilities,
>
> $$P(A_1, A_2, \ldots, A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1, A_2) \cdots P(A_n|A_1, \ldots, A_{n-1}).$$
>
> The commas denote intersections. For example, $P(A_3|A_1, A_2)$ is the probability that A_3 occurs, given that both A_1 and A_2 occur.

We are now ready to introduce the two main theorems about conditional probability: Bayes' rule and the law of total probability (LOTP).

Theorem 2.3.3 (Bayes' rule).

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

This follows immediately from Theorem 2.3.2, which in turn followed immediately from the definition of conditional probability. Yet Bayes' rule has important implications and applications in probability and statistics, since it is so often necessary to find conditional probabilities, and often $P(B|A)$ is much easier to find directly than $P(A|B)$ (or vice versa).

The *law of total probability (LOTP)* relates conditional probability to unconditional probability. It is essential for fulfilling the promise that conditional probability can be used to decompose complicated probability problems into simpler pieces, and it is often used in tandem with Bayes' rule.

Theorem 2.3.4 (Law of total probability (LOTP)).

Let $A_1, \ldots, A_n$ be a partition of the sample space S (i.e., the A_i are disjoint events and their union is S), with $P(A_i) > 0$ for all i. Then

$$P(B) = \sum_{i=1}^{n} P(B|A_i)P(A_i).$$

Proof

Since the A_i form a partition of S, we can decompose B as

$$B = (B \cap A_1) \cup (B \cap A_2) \cup \cdots \cup (B \cap A_n).$$

This is illustrated in Figure 2.3.5, where we have chopped B into the smaller pieces $B \cap A_1$ through $B \cap A_n$. By the second axiom of probability, because these pieces are disjoint, we can add their probabilities to get $P(B)$:

$$P(B) = P(B \cap A_1) + \cdots + P(B \cap A_n) = P(B|A_1)P(A_1) + \cdots + P(B|A_n)P(A_n).$$

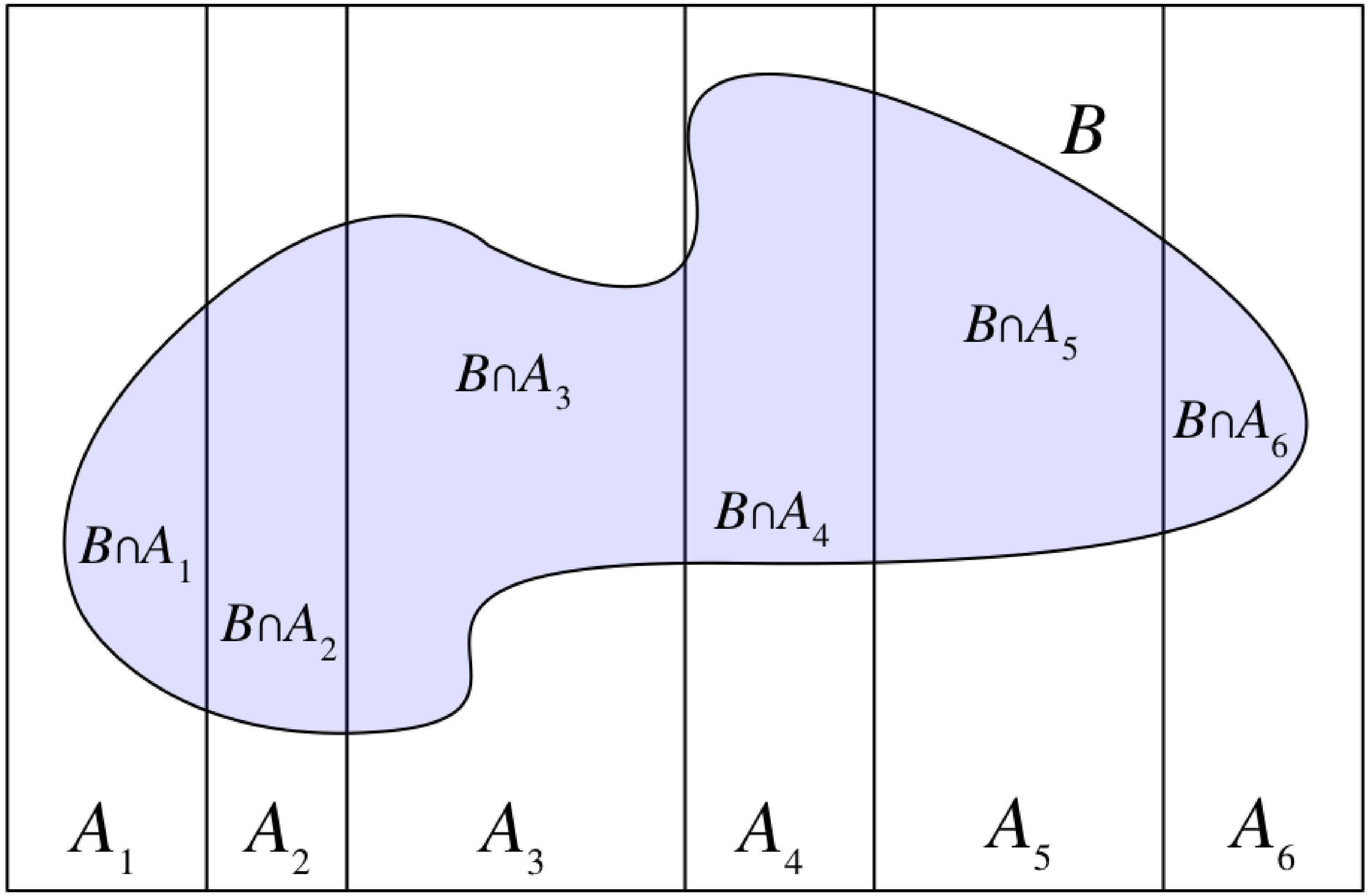

Figure 2.3.5: The A_i partition the sample space; $P(B)$ is equal to $\sum_i P(B \cap A_i)$.

View Larger Image
Image Description

The law of total probability tells us that to get the unconditional probability of B, we can divide the sample space into disjoint slices A_i, find the conditional probability of B within each of the slices, then take a weighted sum of the conditional probabilities, where the weights are the probabilities $P(A_i)$. The choice of how to divide up the sample space is crucial: a well-chosen partition will reduce a complicated problem into simpler pieces, whereas a poorly chosen partition will only exacerbate our problems, requiring us to calculate n difficult probabilities instead of just one!

The next couple examples show how we can use Bayes' rule together with the law of total probability to update our beliefs based on observed evidence.

Example 2.3.6 (Random coin).

You have one fair coin, and one biased coin which lands Heads with probability 3/4. You pick one of the coins at random and flip it three times. It lands Heads all three times. Given this information, what is the probability that the coin you picked is the fair one?

Solution

Let A be the event that the chosen coin lands Heads three times and let F be the event that we picked the fair coin. We are interested in $P(F|A)$, but it is easier to find $P(A|F)$ and $P(A|F^c)$ since it helps to know which coin we have; this suggests using Bayes' rule and the law of total probability. Doing so, we have

$$\begin{aligned} P(F|A) &= \frac{P(A|F)P(F)}{P(A)} \\ &= \frac{P(A|F)P(F)}{P(A|F)P(F) + P(A|F^c)P(F^c)} \\ &= \frac{(1/2)^3 \cdot 1/2}{(1/2)^3 \cdot 1/2 + (3/4)^3 \cdot 1/2} \\ &\approx 0.23. \end{aligned}$$

Example 2.3.7 (Testing for a rare disease).

Jimmy is being tested for a disease called conditionitis, a medical condition that afflicts 1% of the population. The test result is positive, i.e., the test claims that Jimmy has the disease. Let D be the event that Jimmy has the disease and T be the event that he tests positive.

Suppose that the test is "95% accurate"; there are different measures of the accuracy of a test, but in this problem it is assumed to mean that $P(T|D) = 0.95$ and $P(T^c|D^c) = 0.95$. The quantity $P(T|D)$ is known as the *sensitivity* or *true positive rate* of the test, and $P(T^c|D^c)$ is known as the *specificity* or *true negative rate*.

Find the conditional probability that Jimmy has conditionitis, given the evidence provided by the test result.

Solution

Applying Bayes' rule and the law of total probability, we have

$$\begin{aligned} P(D|T) &= \frac{P(T|D)P(D)}{P(T)} \\ &= \frac{P(T|D)P(D)}{P(T|D)P(D) + P(T|D^c)P(D^c)} \\ &= \frac{0.95 \cdot 0.01}{0.95 \cdot 0.01 + 0.05 \cdot 0.99} \\ &\approx 0.16. \end{aligned}$$

So there is only a 16% chance that Jimmy has conditionitis, given that he tested positive, even though the test seems to be quite reliable!

Many people find it surprising that the conditional probability of having the disease given a positive test result is only 16%, even though the test is 95% accurate. The key to understanding this surprisingly high posterior probability is to realize that there are two factors at play: the evidence from the test, and our *prior information* about the prevalence of the disease. The conditional probability $P(D|T)$ reflects a balance between these two factors, appropriately weighing the rarity of the disease against the rarity of a mistaken test result.