LECTURE: BAYES' RULE

1. Bayes' Rule

This rule allows us to calculate $P(A|B)$ knowing $P(B|A)$

Bayes' Rule:

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

2. Law of Total Probability

This is often used in conjunction with the *Law of Total Probability*, which we state below. First we need to define a *partition*:

Definition:

A **partition** of a sample space S is a collection of disjoint subsets whose union is S.

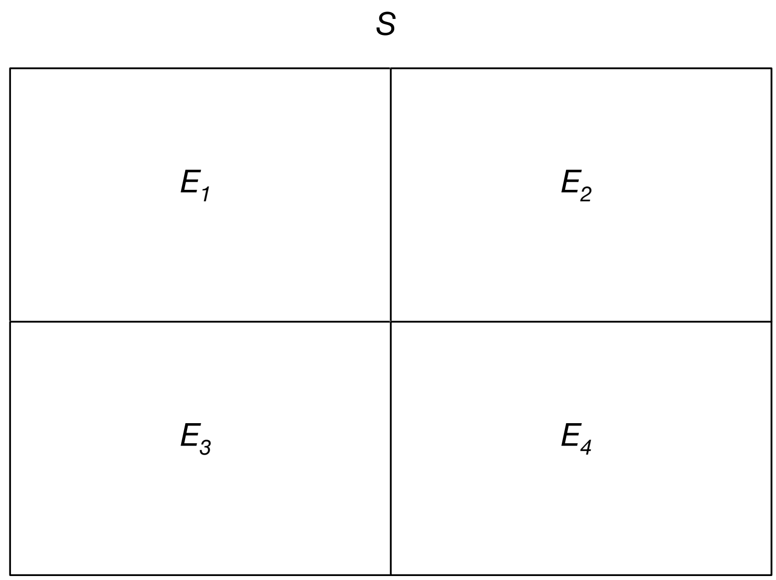
More precisely, it is a collection $\{E_1, E_2, \dots, E_k\}$ of subsets of S such that:

(1) $S = E_1 \cup E_2 \cup \cdots \cup E_k$

(2) $E_i \cap E_j = \emptyset$ for $i \neq j$

Intuitively, think of a partition as a division of S into separate pieces, like dividing up the world into different countries.

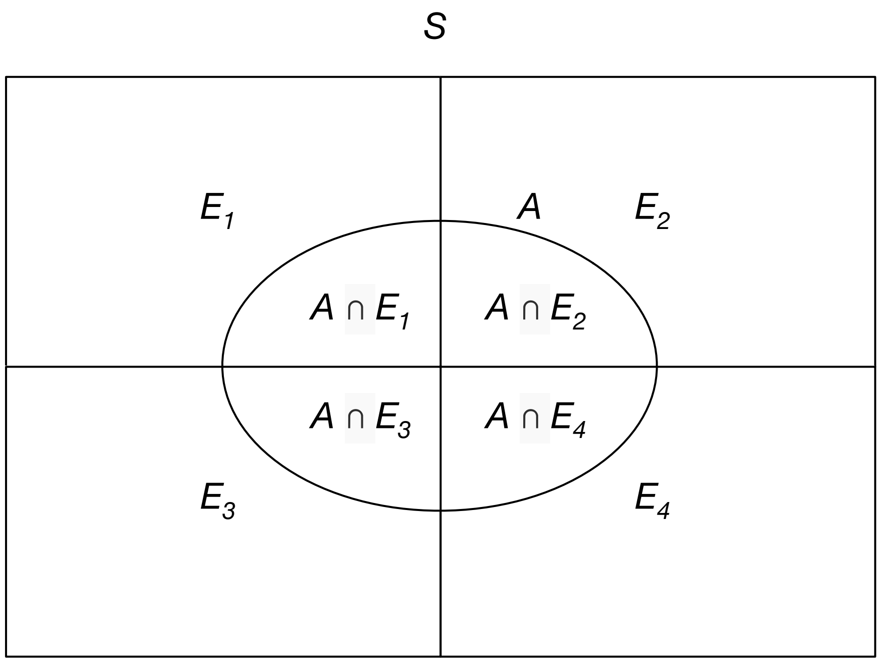
Here is for example a partition of S into 4 subsets:



If $\{E_1, E_2, \ldots, E_k\}$ is a partition of S, and A is any event, then we can decompose A as

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_k)$$

Where this union consists of disjoint sets. For $k = 4$, this is illustrated graphically below:



Law of Total Probability:

Let $\{E_1, E_2, \ldots, E_k\}$ be a partition of S and A is any event, then

$$P(A) = \sum_{i=1}^{k} P(A|E_i)P(E_i)$$

Why? As before, decompose our event A into disjoint sets:

$$A = (A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_k)$$

Then we can write:

$$\begin{aligned} P(A) &= P((A \cap E_1) \cup (A \cap E_2) \cup \cdots \cup (A \cap E_k)) \\ &= P(A \cap E_1) + P(A \cap E_2) + \cdots + P(A \cap E_k) \\ &= P(A|E_1)P(E_1) + P(A|E_2)P(E_2) + \cdots + P(A|E_k)P(E_k) \end{aligned}$$

Where we have used the fact that the decomposition of A is a union of disjoint sets and the definition of conditional probability.

3. General Bayes' Rule

General Bayes' Rule:

Let A and B be two events, and let $\{E_1, E_2, \ldots, E_k\}$ be a partition of S, then

$$P(A|B) = \frac{P(B|A)P(A)}{\sum_{i=1}^{k} P(B|E_i)P(E_i)}$$

Why? This follows from taking Bayes' rule

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

And then expanding $P(B)$ using the Law of Total Probability.

Example 1:

You have two urns of balls. The first urn contains one red ball and three white balls. The second urn contains two red balls and two white balls. You choose an urn at random and then draw a ball from the chosen urn.

What is the probability that you choose the Urn 1 in the first step given that the ball drawn is red?

We are once again looking for $P(B_1|R)$ (notation is the same as before)

Here our partition is $\{B_1, B_2\}$ because balls are either in Urn 1 or in Urn 2 but not both

Writing out Bayes' rule, we get:

$$P(B_1|R) = \frac{P(R|B_1)P(B_1)}{P(R|B_1)P(B_1) + P(R|B_2)P(B_2)}$$

$$P(B_1) = P(B_2) = 1/2$$

$$P(R|B_1) = 1/4$$

$$P(R|B_2) = 1/2$$

So we just plug these in to get:

$$P(B_1|R) = \frac{(1/4)(1/2)}{(1/4)(1/2) + (1/2)(1/2)} = \frac{1/8}{3/8} = \frac{1}{3}$$

which agrees with what we had before!

Example 2:

Consider the following card game. We have 3 cards, colored as follows:

(1) Both sides are red (RR)

(2) Both sides are black (BB)

(3) One side is red and the other side is black (RB)

The three cards are shuffled and one is chosen at random and put down on a table. If the face-up side of the chosen card is red, what is the probability that the other side is black?

Here $\{RR, BB, RB\}$ is a partition of our sample space.

Let R be the event that the face-up side of the chosen card is red. We are interested in $P(RB|R)$.

Then by the general Bayes' Rule, we have

$$\begin{aligned} P(RB|R) &= \frac{P(R|RB)P(RB)}{P(R|RR)P(RR) + P(R|BB)P(BB) + P(R|RB)P(RB)} \\ &= \frac{(1/2)(1/3)}{(1)(1/3) + 0(1/3) + (1/2)(1/3)} \\ &= \frac{1/6}{1/2} \\ &= \frac{1}{3} \end{aligned}$$

Thus the probability that the other side is black is only $1/3$.

Wrong Way: It is tempting to guess that the probability is $1/2$ since there are two possible cards it could be (all red and red-black), and they should be equally likely. It turns out that this is not the case!

Here is one way to see this. Instead of considering three cards, let's think of them as six sides. We can label the six sides as:

(1) R_1 and R_2: the two red sides of the all-red card

(2) B_1 and B_2: the two black sides of the all-black card

(3) R_3 and B_3: the red and black sides of the red-black card

If red is face up, it must be R_1, R_2, or R_3, and these are equally likely. For R_1 and R_2, the other side of the card is red. Only for R_3 is the other side black. Thus the probability of the other side of the card being black is 1/3, which agrees with the result from Bayes' theorem.

4. HIV Case Study

Here is another application of Bayes' theorem which concerns medical testing. This is a classic example in the field of public health.

Example 3:

The OraQuick HIV test is a rapid test for HIV, which can give a result in 20 minutes. The test has the following properties[a]:

(1) Probability of a positive test given that you have the disease (sensitivity): 0.996 (99.6%)

(2) Probability of a negative test given that you don't have the disease (specificity): 0.999 (99.9%)

What is probability that a patient has HIV given that the OraQuick test is positive?

[a]The data is from the CDC website

Let T = "the test is positive" and H = "the patient has HIV"

We are interested in $P(H|T)$

By Bayes' rule:

$$P(H|T) = \frac{P(T|H)P(H)}{P(T)}$$

Since we don't know $P(T)$, the probability of a positive test, we will use the total probability with partition $\{H, H^c\}$

$$P(H|T) = \frac{P(T|H)P(H)}{P(T|H)P(H) + P(T|H^c)P(H^c)}$$

We know the following:

$$\begin{aligned} P(T|H) =& 0.996 \text{ (sensitivity)} \\ P(T|H^c) =& 1 - P(T^c|H^c) = 1 - 0.999 = 0.001 \end{aligned}$$

All that remains is $P(H)$, that is the probability that a randomly-selected patient has HIV.

Real-Life Data: In 2012, the CDC estimates that 1.2 million people in the US have HIV, out of a total population of 319 million. That puts the prevalence of HIV at 0.00376.

So we will take $P(H) = 0.00376$ for now, which gives us $P(H^c) = 1 - 0.00376$. Plugging everything into Bayes' theorem:

$$P(H|T) = \frac{(0.996)(0.00376)}{(0.996)(0.00376) + (0.001)(1 - 0.00376)} = 0.79$$

So we get a positive test in 80% of the cases

Isn't that surprising? Even though the test *seems* very accurate at first, it's actually not. The reason for this is that $P(H)$ is very small, so even the most sensitive and specific test does a poor job of detecting rare diseases.

Variation: In fact, let's repeat this with a different value for the prevalence of HIV. In Fulton County, GA (which contains Atlanta) the prevalence of HIV is approximately 0.0123 (significantly higher than for the nation as a whole). This gives

$$P(H|T) = \frac{(0.996)(0.0123)}{(0.996)(0.0123) + (0.001)(1 - 0.0123)} = 0.925 \text{ which is higher}$$