Topic 6

Conditional Probability and Independence

Bayes Formula and Independence

Outline

Bayes Formula

- Let A be the event that an individual tests positive for some disease and
- let C be the event that the person has the disease.
- We can perform clinical trials to estimate the probability that a randomly chosen individual tests positive given that they have the disease,

$$P\{\text{tests positive}|\text{has the disease}\} = P(A|C),$$

  by taking individuals with the disease and applying the test.
- We would like to use the test as a method of *diagnosis* of the disease. In other words, we would like to give the test and assert the chance that the person has the disease. That is, we want to know the probability with the reverse conditioning

$$P\{\text{has the disease}|\text{tests positive}\} = P(C|A).$$

Bayes Formula

The Public Health Department gives us the following information.

- The test yields a positive result 90% of the time when the disease is present.
- The test yields a positive result 1% of the time when the disease is *not* present.
- One person in 1,000 has the disease.

Let's first think about this intuitively to find the probability

$$P(C|A) = P\{\text{has the disease}|\text{tests positive}\}.$$

- In a city with a population of 1 million people, on average,
  1,000 have the disease and 999,000 do not
- Of the 1,000 that have the disease, on average,
  900 test positive and 100 test negative
- Of the 999,000 that do not have the disease, on average,
  $999{,}000 \times 0.01 = 9990$ test positive and 989,010 test negative.

Bayes Formula

Consequently, among those that test positive, the *odds* of having the disease is

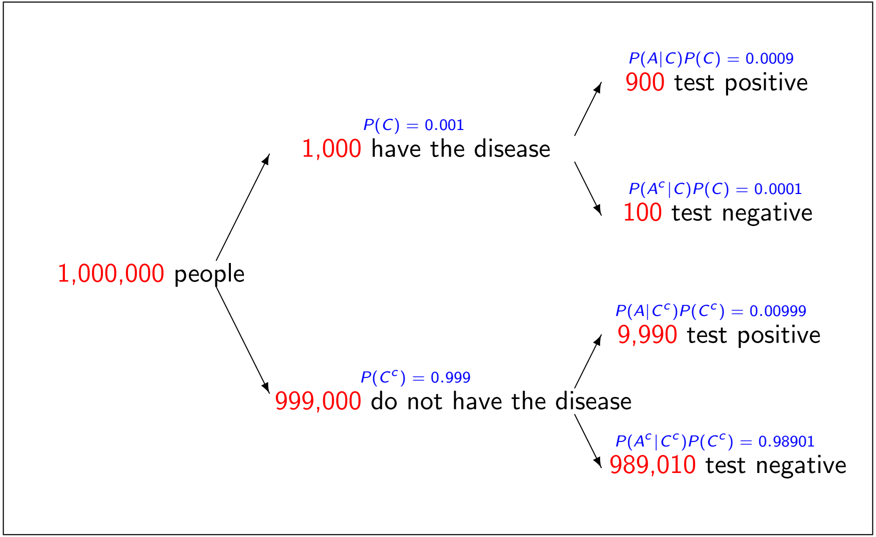
$$\#(\text{have the disease}):\#(\text{does not have the disease})$$

$$900:9990$$

and converting odds to probability we see that

$$P\{\text{have the disease}|\text{test is positive}\} = \frac{900}{900+9990} = 0.0826$$

Tree Diagrams



Bayes Formula

Consequently, among those that test positive, the *odds* of having the disease is

$$\#(\text{have the disease}):\#(\text{does not have the disease})$$

$$900:9990$$
$$P(A|C)P(C):P(A|C^c)P(C^c):$$
$$0.0009:0.999$$

and converting odds to probability we see that

$$P(C|A) = P\{\text{have the disease}|\text{test is positive}\} = \frac{900}{900+9990} = 0.0826$$

$$P(C|A) = \frac{P(A|C)P(C)}{P(A|C)P(C)+P(A|C^c)P(C^c)} = \frac{0.0009}{0.0009+0.0999} = 0.0826$$

Bayes Formula

Let's check the formula

$$
\begin{aligned}
P(C|A) &= \frac{P(A|C)P(C)}{P(A|C)P(C)+P(A|C^c)P(C^c)} && \text{(Bayes Formula)} \\
&= \frac{P(A|C)P(C)}{P(A)} && \text{(Law of Total Probability)} \\
&= \frac{P(A \cap C)}{P(A)} && \text{(Multiplication Principle)}
\end{aligned}
$$

which is just the definition of conditional probability.

Bayes formula can be generalized to the case of a partition $\{C_1, C_2, \ldots, C_n\}$ of Ω chosen so that all the $P(C_i) > 0$. Then, for any event A and any j

$$P(C_j|A) = \frac{P(A|C_j)P(C_j)}{\sum_{i=1}^{n} P(A|C_i)P(C_i)}.$$

Weighing the Odds

The answer may be surprising. Only 8% of those who test positive actually have the disease. This example underscores the fact that good predictions based on intuition can be hard to make.

To determine the probability $P(C|A)$, we must weigh the odds of two terms, each of them itself a product.

- $P(A|C)P(C)$, a big number (the true positive probability) times a small number (the probability of having the disease) *versus*
- $P(A|C^c)P(C^c)$, a small number (the false positive probability) times a large number (the probability of being disease free).

Bayes Formula

Exercise. Fill in the table of values for $P(C|A) = P\{\text{has the disease}|\text{tests positive}\}$.

		percent with disease			
		0.1%		0.5%	
false positive		1%	5%	1%	5%
true positive	90%	0.0826			
	95%				

Which has a bigger impact - a change in the false positive rate or the true positive rate? Give an intuitive explanation for your answer?

Bayes Formula

A box has a 2-headed coin and a fair coin. It is flipped n times, yielding heads each time. What is the probability that the 2-headed coin is chosen? To solve this, note that

$$P\{\text{2-headed coin}\} = \frac{1}{2}, \quad P\{\text{fair coin}\} = \frac{1}{2}.$$

and

$$P\{n \text{ heads}|\text{2-headed coin}\} = 1, \quad P\{n \text{ heads}|\text{fair coin}\} = 2^{-n}.$$

By the law of total probability, $P\{n \text{ heads}\}$

$$\begin{aligned} &= P\{n \text{ heads}|\text{2-headed coin}\}P\{\text{2-headed coin}\} \\ &\quad + P\{n \text{ heads}|\text{fair coin}\}P\{\text{fair coin}\} \\ &= 1 \cdot \frac{1}{2} + 2^{-n} \cdot \frac{1}{2} = \frac{2^n + 1}{2^{n+1}}. \end{aligned}$$

Next, we use Bayes formula. $P\{\text{2-headed coin}|n \text{ heads}\}$

$$= \frac{P\{n \text{ heads}|\text{2-headed coin}\}P\{\text{2-headed coin}\}}{P\{n \text{ heads}\}} = \frac{1 \cdot (1/2)}{(2^n + 1)/2^{n+1}} = \frac{2^n}{2^n + 1} < 1.$$

Bayes Formula

This simple and seemingly silly example is mathematically identical to a question in the vertical transmission of a genetic disease.

- A female knows that she has a history of a allele on her X chromosome for a recessive genetic condition.
- She does not have the condition. So, she knows that she cannot be homozygous for the recessive allele. Consequently, she wants to know her chance of being
  - a carrier (heteorzygous for a recessive allele) or
  - not a carrier (homozygous for the common genetic type).
- The female is a mother with n sons, none of which show the recessive allele on their single X chromosome and so do not have the condition.

Bayes Formula

A box has a 2-headed coin and a fair coin. It is flipped n times, yielding heads each time. What is the probability that the 2-headed coin is chosen? To solve this, note that

$$P\{\text{2-headed coin}\} = \frac{1}{2}, \quad P\{\text{fair coin}\} = \frac{1}{2}.$$

and

$$P\{n \text{ heads}|\text{2-headed coin}\} = 1, \quad P\{n \text{ heads}|\text{fair coin}\} = 2^{-n}.$$

By the law of total probability, $P\{n \text{ heads}\}$

$$\begin{aligned} &= P\{n \text{ heads}|\text{2-headed coin}\}P\{\text{2-headed coin}\} \\ &\quad + P\{n \text{ heads}|\text{fair coin}\}P\{\text{fair coin}\} \\ &= 1 \cdot \frac{1}{2} + 2^{-n} \cdot \frac{1}{2} = \frac{2^n + 1}{2^{n+1}}. \end{aligned}$$

Next, we use Bayes formula.. $P\{\text{2-headed coin}|n \text{ heads}\}$

$$= \frac{P\{n \text{ heads}|\text{2-headed coin}\}P\{\text{2-headed coin}\}}{P\{n \text{ heads}\}} = \frac{1 \cdot (1/2)}{(2^n + 1)/2^{n+1}} = \frac{2^n}{2^n + 1} < 1.$$

Bayes Formula

What is the probability that the female is not a carrier? Let's look at the computation above again. Based on her pedigree, the female estimates that

$$P\{\text{mother not a carrier}\} = p, \quad P\{\text{mother a carrier}\} = 1 - p.$$

and

$$P\{n \text{ sons condition free}|\text{mother not a carrier}\} = 1, \; P\{n \text{ sons condition free}|\text{mother a carrier}\} = 2^{-n}.$$

By the law of total probability, $P\{n \text{ sons condition free}\}$

$$\begin{aligned} &= P\{n \text{ sons condition free}|\text{mother not a carrier}\}P\{\text{mother is not a carrier}\} \\ &\quad + P\{n \text{ sons condition free}|\text{mother a carrier}\}P\{\text{mother a carrier}\} \\ &= 1 \cdot p + 2^{-n} \cdot (1 - p). \end{aligned}$$

Next, we use Bayes formula. $P\{\text{mother not a carrier}|n \text{ sons condition free}\}$

$$= \frac{P\{n \text{ sons condition free}|\text{mother not a carrier}\}P\{\text{mother not a carrier}\}}{P\{n \text{ sons condition free}\}} = \frac{p}{p + 2^{-n}(1 - p)}.$$

Independence

An event A is independent of B if the occurrence of B does not alter the probability of A:

$$P(A) = P(A|B).$$

Multiply this equation by $P(B)$ and use the multiplication rule to obtain

$$P(A)P(B) = P(A|B)P(B) = P(A \cap B).$$

The formula $P(A)P(B) = P(A \cap B)$ is symmetric in A and B. Consequently, we say that A and B are independent.

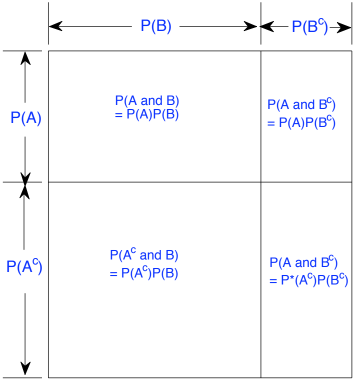


Figure: The Venn diagram for independent events.

Independence

The events $A_1, \cdots, A_n$ are called independent if for any choice $A_{i_1}, A_{i_2}, \cdots, A_{i_k}$ taken from this collection of n events, then

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k}).$$

A similar product formula holds if some of the events are replaced by their complement.

Exercise. Flip 10 biased coins. Their outcomes are independent with the i-th coin turning up heads with probability p_i. Find

$$P\{\text{first coin heads, third coin tails, seventh \& ninth coin heads}\}$$

and

$$P\{\text{fith coin heads, seventh \& eighth coins agree}\}.$$