Module 8.5: Conditional Probability

MCIT Online - CIT592 - Professor Val Tannen

Lecture Notes

Penn Engineering

Conditional probability

Given a probability space (Ω, P) we are interested in an event E in a context in which we already know (for sure!) that another event, U has happened (or is happening).

We write this as $E \mid U$ and we read it as E **conditioned** on U.

The **conditional probability** that we assign to this situation is denoted and defined by

$$\Pr[E|U] = \frac{\sum_{w \in E \wedge w \in U} \Pr[w]}{\sum_{w \in U} \Pr[w]} = \frac{\Pr[E \cap U]}{\Pr[U]} \qquad \text{(provided } \Pr[U] \neq 0\text{)}$$

When $\Pr[U] = 0$ the conditional probability $\Pr[E|U]$ is **undefined**.

One heads or two? (I)

Problem. Alice flips a fair coin twice, independently. Bob did not see the flips but Alice tells him (variant 1) that the **first flip** was heads and asks him the probability that **both** flips were heads. Bob correctly answers $1/2$ (why?). But what if Alice tells him (variant 2) that **at least one of the flips** was heads and asks him the probability that both were heads?

Answer. In variant 1 Bob can simply reason that two heads show iff the second flip is heads. By independence, this happens with probability $1/2$.

For variant 2, Bob might be tempted to answer $1/2$ as well, as the probability of the **other** flip also being heads. This would be OK if Bob knew that the other flip is, say, the second one, as we saw in variant 1. But Bob does not know that.

One heads or two? (II)

Answer (continued). In effect, in both variants Bob is asked to compute a **conditional probability**.

Let E be the event "both flips are heads", F be the event "first flip is heads", and let G be the event "at least one of the flips is heads", that is, $\{\mathtt{HH}, \mathtt{HT}, \mathtt{TH}\}$.

In variant 1 Bob is asked to calculate

$\Pr[E|F] = \Pr[E \cap F]/\Pr[F] = (1/4)/(1/2) = 1/2$

Bob's reasoning led to the correct result.

In variant 2 Bob is asked to calculate

$\Pr[E|G] = \Pr[E \cap G]/\Pr[G] = (1/4)/(3/4) = 1/3$

This is **different** from the $1/2$ that Bob might impetuously answer.

Testing for a rare disease I

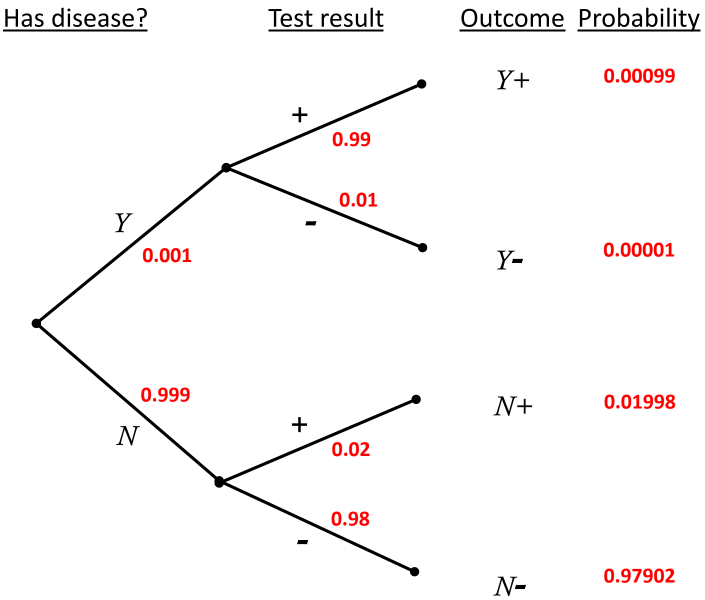
Problem. A test for a rare disease that affects 0.1% of the population is 99% effective on people with the disease (i.e., it gives a **false negative** with probability 0.01). On people who do not suffer from the disease the test gives a **false positive** with probability 0.02. What is the probability that someone who tests positive does, in fact, have the disease?

Answer. Given a person, V, we have two sources of randomness: V has the disease, Y(es) or N(o), and V tests positive $+$ or negative $-$.

Thus we have four outcomes, which we denote $Y+, Y-, N+, N-$ and correspondingly we have four events: $Y = \{Y+, Y-\}$, $T_+ = \{Y+, N+\}$, $N = \{N+, N-\}$, $T_- = \{Y-, N-\}$. Clearly, they are **not** independent!

We use the false negative and false positive statistics as estimates for conditional probabilities: $\Pr[T_-|Y] = 0.01$, $\Pr[T_+|N] = 0.02$.

Testing for a rare disease II



Testing for a rare disease III

Answer (continued). Let's first show that multiplying along the branches is justified by the conditional probability definition:

$\mathrm{Pr}[T_-|Y] = \mathrm{Pr}[T_- \cap Y]/\mathrm{Pr}[Y]$ therefore $\mathrm{Pr}[T_- \cap Y] = \mathrm{Pr}[Y] \cdot \mathrm{Pr}[T_-|Y]$

So $\mathrm{Pr}[Y-] = \mathrm{Pr}[T_- \cap Y] = \mathrm{Pr}[Y] \cdot \mathrm{Pr}[T_-|Y] = 0.001 \cdot 0.01 = 0.00001$

Back to the problem. It asks for the probability that V actually has the disease, knowing that V tested positive. A conditional probability!

$$\begin{aligned}\mathrm{Pr}[Y|T_+] &= \mathrm{Pr}[Y \cap T_+]/\,\mathsf{Pr}[T_+] = \mathrm{Pr}[Y+]/(\mathrm{Pr}[Y+] + \mathrm{Pr}[N+]) \\ &= (0.00099)/(0.00099 + 0.01998) = 0.0472\end{aligned}$$

How long will your battery last?

Problem. Consumers have reported that 80% of new car batteries still work after 10,000 miles and that 40% of them still work after 20,000 miles. If your car battery still works after 10,000 miles, what is the probability that it will last another 10,000 miles?

Answer. Consider the events $L_1 =$ "battery still works after 10,000 mi" and $L_2 =$ "battery still works after 20,000 mi".

Based on the consumer reports we estimate $\Pr[L_1] = 0.8$ and $\Pr[L_2] = 0.4$.

The problem asks for the conditional probability $\Pr[L_2 | L_1]$.

Note that $L_2 \subseteq L_1$ therefore $L_2 \cap L_1 = L_2$.

So $\Pr[L_2|L_1] = \Pr[L_2 \cap L_1] / \Pr[L_1] = \Pr[L_2] / \Pr[L_1] = 0.4/0.8 = 0.5$.

Penn Engineering

Activity : A tree for car batteries

In this activity we will set up a tree of all possibilities for a problem we just solved with another method that did not require us to describe the underlying probability space. We repeat the problem here.

Problem. Consumers have reported that 80% of new car batteries still work after 10,000 miles and that 40% of them still work after 20,000 miles. If your car battery still works after 10,000 miles, what is the probability that it will last another 10,000 miles?

Question. How many outcomes do you think we will use in the probability space?

In the video, there is a box here for learners to put in an answer to the question above. As you read these notes, try it yourself using pen and paper!

Activity : A tree for car batteries (continued)

Answer. There are **three** outcomes, corresponding to the following:

(w_1) The battery still works after 20,000mi.

(w_2) The battery did not last 20,000mi but it still worked after 10,000mi.

(w_3) The battery did not last 10,000mi.

This is because we identify the following two sources of randomness. The first is whether the battery lasts 10,000mi or not. The second only affects the battery life after it has lasted 10,000mi and it is whether it will last another 10,000mi or not.

Activity : A tree for car batteries (continued)

Now consider the events $L_1 =$ "battery still works after 10,000mi" and $L_2 =$ "battery still works after 20,000mi".

In terms of the outcomes we have identified we have $L_1 = \{w_1, w_2\}$ and $L_2 = \{w_1\}$.

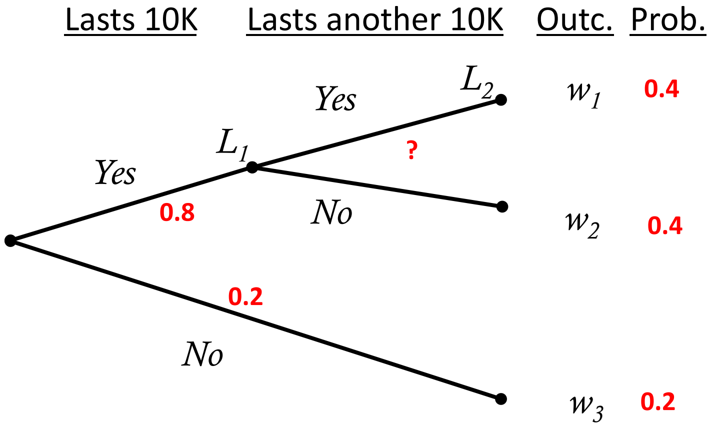
The problem tells us that $\Pr[L_1] = 0.8$ and $\Pr[L_2] = 0.4$.

Therefore we know that $\Pr[w_1] = 0.4$. We can then calculate from $\Pr[L_1] = \Pr[w_1] + \Pr[w_2]$ that $\Pr[w_2] = 0.8 - 0.4 = 0.4$. Also, $\Pr[w_3] = 1 - (\Pr[w_1] + \Pr[w_2]) = 1 - 0.8 = 0.2$.

This leads to the following tree of all possibilities

Penn Engineering

ACTIVITY : A tree for car batteries (continued)



Note that the problem does not give probabilities for the branches that split out of L_1. In fact, it **asks** for one of them (where we put a question mark).

ACTIVITY : A tree for car batteries (continued)

However, recalling how we computed outcome probabilities in the tree of all possibilities for the Monty Hall problem we can recover the answer.

Let p be the probability that labels the Yes branch out of L_1 (the one where we put a question mark in the tree).

Since the outcome probabilities are computed by multiplying along branches we must have $(0.8) \cdot p = 0.4$. Hence $p = (0.4)/(0.8) = 0.5$.

The problem asked: If your car battery still works after 10,000 miles, what is the probability that it will last another 10,000 miles?

That's exactly p and the answer is 0.5 just as in the first solution.