

# Probability and Expected Value

Brian Weatherson, February 2, 2018

## Inverting Conditional Probability

Here is a common enough situation. We have two variables,  $X, Y$  that we initially don't know the values of. We know, however, that they are (in some informal sense) correlated. We know the prior probability for each value of  $X$ , usually from frequency data. We also know the probability for each value of  $Y$  given a value of  $X$ . We want to solve for the conditional probability of a value of  $X$  given a value of  $Y$ . Or, equivalently, we want to know what the probability of a value of  $X$  should be once we learn a value of  $Y$ .

Let's say the possible values of  $X$  are  $x_1, \dots, x_m$  and the possible values of  $Y$  are  $y_1, \dots, y_n$ . For each  $x_i$  the prior probability that  $X = x_i$  is  $c_i$ . And for each pair  $x_i, y_j$  the probability that  $Y = y_j$  given  $X = x_i$  is  $p_{ij}$ . Then we can make a table up with the values of  $X$  on the rows, and the values of  $Y$  on the columns, and each cell representing the (prior) probability of a pair of  $X$  value and  $Y$  value. (When I say a table, I mean this somewhat literally; I usually do this kind of computation in Excel or some other spreadsheet program.)

|         | $y_1$        | $y_2$        | $\dots$ | $y_j$        | $\dots$ | $y_n$        |
|---------|--------------|--------------|---------|--------------|---------|--------------|
| $x_1$   | $c_1 p_{11}$ | $c_1 p_{12}$ | $\dots$ | $c_1 p_{1j}$ | $\dots$ | $c_1 p_{1n}$ |
| $x_2$   | $c_2 p_{21}$ | $c_2 p_{22}$ | $\dots$ | $c_2 p_{2j}$ | $\dots$ | $c_2 p_{2n}$ |
| $\dots$ | $\dots$      | $\dots$      | $\dots$ | $\dots$      | $\dots$ | $\dots$      |
| $x_i$   | $c_i p_{i1}$ | $c_i p_{i2}$ | $\dots$ | $c_i p_{ij}$ | $\dots$ | $c_i p_{in}$ |
| $\dots$ | $\dots$      | $\dots$      | $\dots$ | $\dots$      | $\dots$ | $\dots$      |
| $x_m$   | $c_m p_{m1}$ | $c_m p_{m2}$ | $\dots$ | $c_m p_{mj}$ | $\dots$ | $c_m p_{mn}$ |

The probability of each  $Y$  value is then given by adding up the numbers in each column. So the probability of  $Y = y_j$  is given by this formula:

$$c_1 p_{1j} + c_2 p_{2j} + \dots + c_i p_{ij} + \dots + c_m p_{mj}$$

And then the probability of  $X = x_i$  conditional on  $Y = y_j$  is the probability of  $X = x_i \wedge Y = y_j$  divided by the probability of  $Y = y_j$ . That is, it is

$$\frac{c_i p_{ij}}{c_1 p_{1j} + c_2 p_{2j} + \dots + c_i p_{ij} + \dots + c_m p_{mj}}$$

That's the full theory, though it might be good to work through an example or two. Here's one that's slightly more complicated than on Tuesday, or in the book.

There are only two status we care about: having disease D or not having it. (Those are the X's.) The test, however, has three outcomes: positive, unclear, or negative. Here are the relevant probabilities:

- Prior probability of having the disease is 7%, i.e., 0.07.
- Probability of positive test given the disease is 80%, and probability of an unclear test is 20%.
- Probability of negative test given not having the disease is 85%, and probability of an unclear test is 10%.

So we can deduce two more relevant acts

- Prior probability of not having the disease is 0.93, i.e.,  $1 - 0.07$ .
- Probability of positive test given not having the disease is 0.05, i.e.,  $1 - 0.85 - 0.1$ .

So now we can make the full table.

|             | Positive Test               | Unclear Test              | Negative Test               |
|-------------|-----------------------------|---------------------------|-----------------------------|
| Has Disease | $0.07 \times 0.8 = 0.056$   | $0.07 \times 0.2 = 0.014$ | $0.07 \times 0 = 0$         |
| No Disease  | $0.93 \times 0.05 = 0.0465$ | $0.93 \times 0.1 = 0.093$ | $0.93 \times 0.85 = 0.7905$ |

The probability of having the disease given a positive test is

$$\frac{0.056}{0.056 + 0.0465} \approx 0.546$$

The probability of having the disease given an unclear test is

$$\frac{0.014}{0.014 + 0.093} \approx 0.13$$

And obviously the probability of having the disease given a negative test is 0.

## Random Variables

A **random variable** is simply a variable that takes different numerical values in different states. In other words, it is a function from possibilities to numbers. It need not be ‘random’ in any familiar sense. The function from possible situations to the value of  $2 + 2$  in that situation is a random variable, albeit a constant one. It’s just a slightly confusing term for any variable that takes different, numerical, values in different situations.

Typically, random variables are denoted by capital letters. So we might have a random variable  $X$  whose value is the age of the next President of the United States, at his or her inauguration. Or we might have a random variable  $Y$  that is the number of children you will have in your lifetime. Basically any mapping from possibilities to numbers can be a random variable.

Here’s one topical example. You’ve asked each of your friends who will win the big football game this weekend. 9 said the home team will win. (I don’t know who is the home team in the Superbowl; I think it’s the AFC team.) 5 said the away team will win. Then we can let  $X$  be a random variable measuring the number of your friends who correctly predicted the result of the game.

$$X = \begin{cases} 9, & \text{if the home team wins,} \\ 5, & \text{if the away team wins.} \end{cases}$$

## Expected Value

Given a random variable  $X$  and a probability function  $\text{Pr}$ , we can work out the **expected value** of that random variable with respect to that probability function. Intuitively, the expected value of  $X$  is a weighted average of the possible values of  $X$ , where the weights are given by the probability (according to  $\text{Pr}$ ) of each value coming about.

More formally, we work out the expected value of  $X$  this way.

- For each possibility, we multiply the value of  $X$  in that case by the probability of the possibility obtaining.
- Then we sum the numbers we've produced, and the result is the expected value of  $X$ .
- We'll write the expected value of  $X$  as  $\text{Exp}(X)$ .

So in the earlier example, if the probability that the home wins is 0.8, and the probability that the away team wins is 0.2, then

$$\begin{aligned}\text{Exp}(X) &= 9 \times 0.8 + 5 \times 0.2 \\ &= 7.2 + 1 \\ &= 8.2\end{aligned}$$

The expected value of  $X$  isn't in any sense the value that we expect  $X$  to take. It's more like an average. If this kind of situation recurs a lot, you would expect the long run average value  $X$  takes to be roundabout the expected value. That's a better way of conceptualising what expected values are.

## Maximise Expected Value

Here's the core principle of most modern theories of decision. If you want to promote some value, you should **maximise its expected value**. So if you're a business, and you want to make money, you should maximise the expected profit from ventures. If you're a sport team, and you want to score a lot of points, you should maximise the expected points from each play.

The reason is that in the long run, is what's going to produce the best results. But what if there is no long run? Well, the same reasoning still goes through as long as the thing we are focussing on is the only thing we care about.

And it might not be. Sports teams don't just try to score points, they try to prevent the other team scoring, and ultimately they try to win games. Businesses don't just try to maximise profits; they might have reasons for earning profits at a particular time, or they might have other non-financial goals that they are trying to promote.

## Utility

The orthodox view is that (for rational, coherent) people we can collapse all the things that they care about into a single measure, utility. A rational choice is one that maximises expected utility. But note this is a little misleading; it suggests that what makes a choice rational is that it maximises expected utility. Some theorists think that is true, but others think the order of explanation goes the other way. (I'm sort of on that latter team.)

Anyway, the big picture is that there is some random variable  $U$  that measures how well off you are in each possibility. And the key feature of  $U$  is that it obeys the following two constraints, which we sort of take to amount to the same thing.

If  $U(w_2) - U(w_1) = U(w_3) - U(w_2)$  then

- The amount that  $w_3$  is better than  $w_2$  equals the amount that  $w_2$  is better than  $w_1$ .
- You are indifferent between having  $w_2$  as a sure outcome, and a 50/50 chance of ending up in either  $w_1$  or  $w_3$ .

It's worth thinking through this in some more practical cases. Imagine you are offered a choice of two envelopes, one Red, and one Green. The Red envelope has a lottery ticket in it that will lead to you either winning \$ $R$ , or winning nothing, each of them equally likely. The Green envelope has cash in it, specifically \$ $G$ . For each of the following values for  $R$ , find a value for  $G$  such that you personally would be equally happy to be given the Red envelope or the Green envelope.

| Red             | Green |
|-----------------|-------|
| \$10            | ??    |
| \$100           | ??    |
| \$1,000         | ??    |
| \$10,000        | ??    |
| \$1,000,000     | ??    |
| \$1,000,000,000 | ??    |

Theory is silent on how you should answer this; it depends on how much you value near versus far increases in wealth.