

Conditional Probabilities and Bayes Theorem

Jeffrey Phillips Freeman, John Doe

Abstract

A simple explanation of conditional probabilities.

Keywords

Statistics, Math

Syncleus, NoTech

Corresponding author: jeffrey.freeman@syncleus.com

©April 6, 2017 - April 19, 2017

Contents

1 Probability Space

2

Draft: April 19, 2017

Draft: April 19, 2017

I've been getting a lot of questions from friends lately about what Bayes Theorem means. The confusion is understandable because it appears in a few models that seem to be completely unrelated to each other. For example we have Naive Bayes Classifiers and Bayesian Networks which operate on completely different principles. Moreover this is compounded with a lack of understanding regarding unconditional and conditional probabilities.

In this article I offer a tutorial to help bring the lay person up to speed with some basic understanding on these concepts and how Bayes Theorem can be applied.

1. Probability Space

We have to start by explaining a few terms and how they are used.

A **Random Trial** is a trial where we perform some random experiment. For example it might be flipping a coin or rolling dice.

The **Sample Space** of a Random Trial, typically denoted by Ω , represents all possible outcomes for the Random Trial being performed. So for flipping a coin the outcome can be either heads or tails, so the Sample Space would be a set containing only these two values.

$$\Omega = \{Heads, Tails\}$$

For rolling dice the Sample Space would be the set of all the various faces for the dice being rolled. When rolling only one standard six sided die the Sample space would be as follows.

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

In both of these examples the Random Trial being performed will select an outcome from their respective Sample Space at random. In these trials each outcome has an equal chance of being selected, though that does not necessarily need to be the case.

For our purposes here I want to formulate a Random Trial thought experiment that simulates a medical trial consisting of 10 patients. We will be using this example throughout much of this tutorial. Therefore our Sample Space will be a set consisting of 10 elements, each element represents a single unique patient in the trial. Patients are represented with the variable x with a subscript from 1 to 10 that uniquely identifies each patient.

$$\Omega = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$$

An **Event** in the context of probabilities is a set of outcomes that can be satisfied. Typically these sets are represented using lowercase greek letters such as α or β . For example if we were rolling a single die and wanted it to land on an odd number then the Event representing an odd outcome would be represented as the following set.

$$\alpha = \{1, 3, 5\}$$

Similarly if we simply wanted to roll the number 6 then the Event would be a set containing only that one number.

$$\alpha = \{6\}$$

The **Event Space**, often denoted by \mathcal{F} , is the set of all Events to be observed. It is a set of sets that represents every possible combination of subsets of the Sample Space or some part thereof. Not all Events in the event space need to be possible however. For example if we are talking about flipping a coin then the Event Space would have, at most, 4 members representing outcomes for Heads, Tails, either Heads or Tails, and neither Heads nor Tails. We could represent this with the following set notation.

$$\mathcal{F} = \{\{\}, \{Heads\}, \{Tails\}, \{Heads, Tails\}\}$$

The empty set is usually represented with the \emptyset symbol. So the previous set can be rewritten using this shorthand as follows.

$$\mathcal{F} = \{\emptyset, \{Heads\}, \{Tails\}, \{Heads, Tails\}\}$$

Notice that one of the members of the Event Space is equivalent to the Sample space. the fact that it contains both the empty set and the Sample Space as members is more a matter of mathematical completeness and plays a role in making some mathematical proofs easier to carry out. For our purposes here they will largely be ignored.

At this point I want to go over a little bit of mathematical notation that may help when reading other texts on the subject. The first is the concept of a **Power Set**. The Power Set is simply every possible combination of subsets for a particular set. In the example above regarding the coin toss Event Space we can say that the Event Space specified is the Power Set of the Sample Space. The notation for the Power Set is the number 2 with an exponent that is a set. For example short hand for the above Event Space definition could have been the following.

$$\mathcal{F} = 2^{\Omega}$$

Every Event Space must be either equal to, or a subset of, the Power Set of the Sample Space. We can represent that with the following set notation.

$$\mathcal{F} \subseteq 2^{\Omega}$$

Going back to our example of patients in a clinical trial we might want to know what the chance is of selecting a patient at random that has a fever. In that case the Event would be the set of all patients that have a fever and the outcome would be a single patient selected at random. Each Event is an element in the Event Space. So we will denote it as \mathcal{F} with a subscript so it is easier to read than it would be using arbitrary greek lowercase letters, as is the usual convention. If 3 of the 10 patients in our Sample Space have a fever we can represent the fever Event as follows.

$$\mathcal{F}_{fever} = \{x_1, x_6, x_8\}$$

This means that if we select a patient at random and that patient is a member of the \mathcal{F}_{fever} set then that patient has a fever and the outcome has satisfied the event. Similarly we can define the event representing patients that have the flu with the following notation.

$$\mathcal{F}_{flu} = \{x_2, x_4, x_6, x_8\}$$

As stated earlier Events are simply members of the Event Space. This can be indicated using the following set notation which simply states that the flu Event is a member of the Event Space and the fever Event is also a member of the Event Space.

$$\mathcal{F}_{fever} \in \mathcal{F}$$

$$\mathcal{F}_{flu} \in \mathcal{F}$$

Similarly if we wish to indicate that the fever and flu Events are subsets of the Event Space we can do so using the following notation.

$$\mathcal{F}_{fever} \subset \Omega$$

$$\mathcal{F}_{flu} \subset \Omega$$

The only term left to define is the **Probability Space**. This is just the combination of the Event Space, the Sample Space, as well as the probability of each of the Events taking place. It represents all the information we need to determine the chance of any possible outcome occurring. It is denoted as a 3-tuple containing these three things. The probability, P, represents a function that maps Events in \mathcal{F} to probabilities.

$$(\Omega, \mathcal{F}, P)$$

Unconditional Probability

This is where things get interesting. Since we have all the important terms defined we can start talking about actual probabilities. We start with the simplest type of probability, the **Unconditional Probability**. These are the sort of probability most people are familiar with. It is the chance that an outcome will occur independent of any other Events. For example if I flip a coin I can say the probability of it landing on Heads is 50%; this would be an Unconditional Probability.

If all the outcomes in our Sample Space have the same chance of being selected by a Random Trial then calculating the Unconditional Probability is rather easy. The Event would represent all desired outcomes from our Sample Space. So if we wanted to flip a coin and get heads then our Event is a set with a single member and the Sample Space consists of only two members, the possible outcomes. We can write this as the following.

$$\Omega = \{Heads, Tails\}$$

$$\mathcal{F}_{Heads} = \{Heads\}$$

If the above event is satisfied by the flip of a coin it means the outcome of the coin toss was heads. To calculate the probability for this event we simply count the number of members in the Event Set and divide it by the number of members in the Sample Space. In this case the result is 50% but we can represent that as follows.

$$P(\mathcal{F}_{Heads}) = \frac{1}{2}$$

The number of members in a set is called **Cardinality**. We can represent that using notation that is the same as the absolute value sign used around a set. Therefore we can represent the previous equation using the following notation.

$$P(\mathcal{F}_{Heads}) = \frac{|\mathcal{F}_{Heads}|}{|\Omega|} = \frac{1}{2}$$

We can generalize this for any Event represented as \mathcal{F}_i with the following definition for calculating an Unconditional Probability.

$$P(\mathcal{F}_i) = \frac{|\mathcal{F}_i|}{|\Omega|}$$

Now let's apply this to our clinical trial example from earlier. Say we wanted to calculate the chance of selecting someone from the 10 patients in the trial, at random, such that the person selected has a fever. We can calculate that with the following.

$$P(\mathcal{F}_{fever}) = \frac{|\mathcal{F}_{fever}|}{|\Omega|} = \frac{3}{10}$$

We can also do the same for calculating the chance of randomly selecting a patient that has the flu.

$$P(\mathcal{F}_{flu}) = \frac{|\mathcal{F}_{flu}|}{|\Omega|} = \frac{4}{10} = \frac{2}{5}$$

Conditional Probability

A **Conditional Probability** takes this idea one step further. A Conditional Probability specifies the probability of an event being satisfied if it is known that another event was also satisfied. For example using our clinical trial thought experiment one might ask what is the probability of someone having the flu if we know that person has a fever. This would be represented with the following notation.

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever})$$

Assuming that having a fever has some effect on the likelihood of having the flu then this probability would be different than the chance for just any randomly selected member having the flu, after all people with fevers are more likely to have the flu than people without a fever.

Since we already know which patients have the flu and which have a fever it is easy to determine an answer to this question. To calculate the probability we can look at how many patients in our Sample Space have a fever and what percentage of those patients with fever also have the flu. By looking at the data we can see that there are 3 patients with fevers and of those patients only 2 of them have the flu. So the answer is $\frac{2}{3}$.

$$\mathcal{F}_{fever} = \{x_1, x_6, x_8\}$$

$$\mathcal{F}_{flu} = \{x_2, x_4, x_6, x_8\}$$

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever}) = \frac{2}{3}$$

We can generalize this statement by saying that we take the intersection of the sets that represent the Event for patients with the flu and patients with a fever. The intersection is just the set of all the elements that those two sets have in common.

The symbol for intersection is \cap , therefore we can show the intersection of these two sets as follows.

$$\mathcal{F}_{flu} \cap \mathcal{F}_{fever} = \{x_6, x_8\}$$

Another way to look at calculating the Conditional Probability would be to take the Cardinality of the intersection of these two Events and divide it by the cardinality of the conditional Event that has been satisfied. So now we have the following.

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever}) = \frac{|\mathcal{F}_{flu} \cap \mathcal{F}_{fever}|}{|\mathcal{F}_{fever}|} = \frac{2}{3}$$

We can also ask a similar, but markedly different, question. If we know a patient has the flu what is the chance that same patient will have a fever. For this we can use the same logic as above and come up with the following.

$$P(\mathcal{F}_{fever} | \mathcal{F}_{flu}) = \frac{|\mathcal{F}_{flu} \cap \mathcal{F}_{fever}|}{|\mathcal{F}_{flu}|} = \frac{2}{4} = \frac{1}{2}$$

As you can see the only thing that changed is the denominator which is now the Cardinality of the flu Event rather than the fever Event. We can generalize the equation for calculating a Conditional Probability as follows.

$$P(\mathcal{F}_i | \mathcal{F}_j) = \frac{|\mathcal{F}_i \cap \mathcal{F}_j|}{|\mathcal{F}_j|}$$

Bayes Theorem

Bayes Theorem itself is remarkably simple on the surface yet immensely useful in practice. In its simplest form it lets us calculate a Conditional Probability when we have limited information to work with. If we only knew, for example, the probabilities for $P(F_i | F_j)$, $P(F_i)$, and $P(F_j)$, then using Bayes Theorem we could calculate the probability for $P(F_j | F_i)$. The precise equation for Bayes Theorem is as follows.

$$P(\mathcal{F}_i | \mathcal{F}_j) = \frac{P(\mathcal{F}_j | \mathcal{F}_i) \cdot P(\mathcal{F}_i)}{P(\mathcal{F}_j)}$$

Let's say we didn't know all the details of the clinical trial from earlier; we have no idea what the Sample Space is or what members belong to each Event set. All we know is the probability that someone will have a fever at any given time, the probability they will have the flu, and the probability that someone with the flu has a fever. From this limited information, and using Bayes Theorem it would be possible to infer the probability of having the flu if you have a fever. First let's copy the probabilities we know to match what we previously calculated manually.

$$P(\mathcal{F}_{fever}) = \frac{3}{10}$$

$$P(\mathcal{F}_{flu}) = \frac{2}{5}$$

$$P(\mathcal{F}_{fever} | \mathcal{F}_{flu}) = \frac{1}{2}$$

Using only this information, along with Bayes Theorem, we can calculate the probability of someone having the flu if they have a fever as follows.

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever}) = \frac{P(\mathcal{F}_{fever} | \mathcal{F}_{flu}) \cdot P(\mathcal{F}_{flu})}{P(\mathcal{F}_{fever})}$$

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever}) = \frac{\frac{1}{2} \cdot \frac{2}{5}}{\frac{3}{10}}$$

$$P(\mathcal{F}_{flu} | \mathcal{F}_{fever}) = \frac{2}{3}$$

This solution of course agrees with our earlier results when we were able to calculate the answer by manually counting the data. However, this time we did not have to use the data directly.

Let's do one more example to drive the point home. Say we have a test for Tuberculosis, TB, that is 95% accurate. That is to say that if you have TB then 95% of the time the test will give you a positive result. Similarly if you do not have TB then only 95% of the time will you get a negative result. We can represent this as follows.

$$P(\mathcal{F}_{positive} | \mathcal{F}_{infected}) = \frac{19}{20}$$

Furthermore let's say we know that only one in a thousand members of the population are infected with TB at any one time. We can demonstrate this as follows.

$$P(\mathcal{F}_{infected}) = \frac{1}{1000}$$

Finally let's say when tested on the general population that 509 out of every 10,000 people received a positive result. We can represent that with the following.

$$P(\mathcal{F}_{positive}) = \frac{509}{10000}$$

With this information it is possible to calculate the probability someone will have TB if they receive a positive test result. Using Bayes Theorem we can solve for the probability as follows.

$$P(\mathcal{F}_{infected} | \mathcal{F}_{positive}) = \frac{P(\mathcal{F}_{positive} | \mathcal{F}_{infected}) \cdot P(\mathcal{F}_{infected})}{P(\mathcal{F}_{positive})}$$

$$P(\mathcal{F}_{infected} | \mathcal{F}_{positive}) = \frac{\frac{19}{20} \cdot \frac{1}{1000}}{\frac{509}{10000}}$$

$$P(\mathcal{F}_{infected} | \mathcal{F}_{positive}) = \frac{19}{1018} = 0.018664 = 1.8664\%$$

This gives us a very surprising result. It says that of the people who take the TB test and show up positive less than 2% of them actually have TB. This demonstrates the importance of using very accurate clinical tests when testing for diseases that have a low occurrence in the population. Even a small error in the test can give false positives at an alarmingly high rate. Shelah [1969]

References

Saharon Shelah. Stable theories. *Israel Journal of Mathematics*, 7:187–202, 1969.

Draft: April 19, 2017

Draft: April 19, 2017