

# Conditional Probability

By Declan Stacy

## 1 Introduction

Have you ever been told that the probability of two events occurring is simply the product of the probabilities of each individual event occurring? It sounds reasonable. The probability of flipping a coin twice and getting two heads is  $\frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$  because there is a  $\frac{1}{2}$  chance of flipping heads on the first flip, and a  $\frac{1}{2}$  chance of flipping heads on the second flip. However, it turns out that this statement is not always true.

## 2 Conditional Probability and Independence

What if I wanted to know the probability of flipping three coins and getting at least one tails and at least one heads? By the logic above, the answer would be  $\frac{7}{8} \cdot \frac{7}{8} = \frac{49}{64}$  because there is a  $(1 - \frac{1}{8})$  chance of getting not all heads (aka at least one tails) and a  $(1 - \frac{1}{8})$  chance of getting not all tails (aka at least one heads). But does  $\frac{49}{64}$  make sense? After all, there are only 8 possible outcomes for flipping 3 coins.

The problem is, knowing that there is at least one tails affects the chances of getting at least one heads. You know now that the result could not have been HHH, so your 8 possible outcomes has shrunk to 7. In other words, your sample space (the set of all possible outcomes) has been altered. Now, instead of having a  $\frac{7}{8}$  chance of getting at least one heads, you have a  $\frac{6}{7}$  chance, because there are 7 possible outcomes and 6 of them include a heads. We call this new probability a conditional probability, in this case the probability of flipping at least one heads given that you have flipped at least one tails. We can denote this as  $P(B|A)$ , where B is the event “at least one heads” and A is the event “at least one tails”.

Now that we know why we failed, let's see if we can get the correct answer. What we are looking for is  $P(A \cap B)$  (the probability of A and B). First, we need A to happen, which occurs with probability  $P(A) = \frac{7}{8}$ . After we know that A has happened, we need B to also happen, which occurs with probability  $P(B|A) = \frac{6}{7}$ . These are the two probabilities that must be multiplied (see if you can explain why).

Thus, we can write

$$P(A \cap B) = P(A) \cdot P(B|A) \tag{1}$$

In our case, this is  $\frac{7}{8} \cdot \frac{6}{7} = \frac{3}{4}$ . This makes sense because as long as I don't flip all heads or all tails, I have at least one head and at least one tails. There are  $8 - 2 = 6$  ways to not get those two bad outcomes, so our answer should be  $\frac{6}{8}$ , and it is.

What if we have three events? Or four? In general, for events  $A_1, A_2, \dots, A_n$ , we can write

$$P(A_1 \cap \dots \cap A_n) = P(A_1) \cdot P(A_2|A_1) \cdot P(A_3|A_1 \cap A_2) \cdot \dots \cdot P(A_n|A_1 \cap \dots \cap A_{n-1}) \quad (2)$$

which comes from successive applications of 1.

So, when is the statement that  $P(A \cap B) = P(A) \cdot P(B)$  true? By comparison with 1, this would mean that  $P(B|A) = P(B)$ . In words, this means that knowing that the event A has occurred does not affect the chances of event B occurring. We call this independence. So, the probability of multiple events occurring is the product of the probabilities of each individual event occurring if and only if the events are independent.

We now know the relationship between conditional probability and the probability of multiple events occurring, but how do we calculate these conditional probabilities in the first place? Going back to 1, we can divide both sides by  $P(A)$  (as long as  $P(A)$  is not 0) to obtain

$$\frac{P(A \cap B)}{P(A)} = P(B|A) \quad (3)$$

### 3 Total Probability Theorem

Let's look at another example problem where conditional probability is useful. I flip a coin and roll a die, and the die roll determines how much money I will spend on Purell this week. If the flip is heads, I add 95 to my die roll. If it is tails, I multiply my die roll by 50. What is the probability of me spending at least \$100 on Purell this week? We can split the problem up into two cases:

**Case 1:** I rolled heads

In this case, there are 2 die rolls (5 and 6) that result in me spending at least \$100 out of 6 possible die rolls. Thus, there is a  $\frac{1}{3}$  chance for this case.

**Case 2:** I rolled tails

In this case, there are 5 die rolls (2 through 6) that result in me spending at least \$100 out of 6 possible die rolls. Thus, there is a  $\frac{5}{6}$  chance for this case.

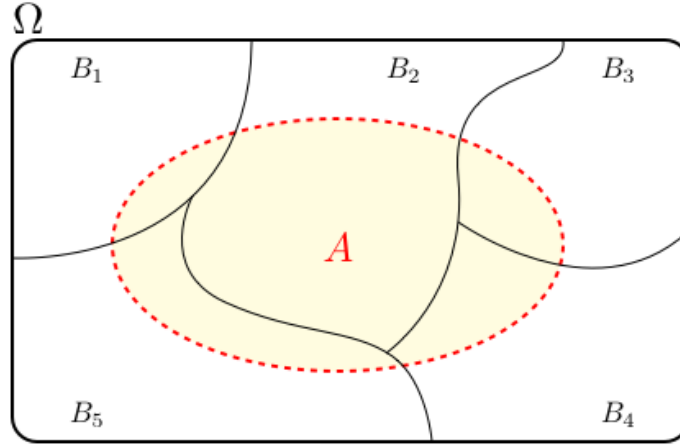
Since case 1 and case 2 each occur with probability  $\frac{1}{2}$ , my total probability is  $\frac{1}{2} \cdot \frac{1}{3} + \frac{1}{2} \cdot \frac{5}{6} = \frac{7}{12}$ . I said conditional probability was useful in this problem. So where did I use it?

Let's call the event of rolling heads H, the event of rolling tails T, and the event of me spending at least \$100 on Purell this week A. The  $\frac{1}{3}$  represents the probability of A given that I rolled heads,  $P(A|H)$ . The  $\frac{5}{6}$  represents the probability of A given that I rolled tails,  $P(A|T)$ . The two  $\frac{1}{2}$ 's represent the probabilities of rolling heads and tails,  $P(H)$  and  $P(T)$ , respectively. To compute  $P(A)$ , I computed  $P(H) \cdot P(A|H) + P(T) \cdot P(A|T)$ . More generally, for disjoint events  $B_1, B_2, \dots, B_n$  such that  $B_1 \cup B_2 \cup \dots \cup B_n$  is the sample space,

$$P(A) = \sum_{i=1}^n P(B_i) \cdot P(A|B_i) = \sum_{i=1}^n P(A \cap B_i) \quad (4)$$

Where the second equality comes from 1. This is called the Total Probability Theorem.

This equation is simply a justification of what you do every time you use casework in a probability problem, but if it is not intuitive, then try to understand this “proof by picture”:



For a formal proof, we must introduce the following axiom (all of probability theory is based on three axioms proposed by Andrey Kolmogorov, and this is one of them):

$$P(A_0 \cup A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) \quad \text{if } A_0, A_1, \dots, A_n \text{ are disjoint events} \quad (5)$$

Since  $B_1, B_2, \dots, B_n$  are disjoint,  $(B_1 \cap A), (B_2 \cap A), \dots, (B_n \cap A)$  are also disjoint. This means that we can apply 5:

$$P((B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_n \cap A)) = \sum_{i=1}^n P(A \cap B_i).$$

Also, since  $B_1 \cup B_2 \cup \dots \cup B_n$  is the sample space, then for an event  $A$  in the sample space,  $A \subset (B_1 \cup B_2 \cup \dots \cup B_n)$ . Thus,  $(B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_n \cap A) = A$  (which is what the picture shows). Substituting  $A$  for  $(B_1 \cap A) \cup (B_2 \cap A) \cup \dots \cup (B_n \cap A)$  in the equation above:

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(A \cap B_i) \\ &= \sum_{i=1}^n P(B_i) \cdot P(A|B_i) \quad (\text{from 1}) \quad \square \end{aligned}$$

There is also an analogous formula for computing expectations with conditional probability, where  $A_1, A_2, \dots, A_n$  are disjoint events such that  $A_1 \cup A_2 \cup \dots \cup A_n$  is the sample space of a random variable  $X$  which will not be proven here:

$$E(X) = \sum_{i=1}^n E(X|A_i) \cdot P(A_i) \quad (6)$$

## 4 Bayes' Theorem

There is another very important application of conditional probability that comes up a lot in everyday life: inference. For example, when we go to the doctor and get tested for COVID-19, we want to know how likely it is that we have the disease given that we tested positive. The question also gives you that 10% of the population has the virus, that the test is 99% accurate when the patient has the virus, and 90% accurate when the patient does not have the virus.

First let's understand what the question gives us. There are two events we are concerned with: you having the virus, and you testing positive. We will call these events A and B, respectively. You are given that  $P(A) = .1$ ,  $P(B|A) = .99$ , and  $P(B|\neg A) = 1 - .9 = .1$  (if the test is 90% accurate when you don't have the virus, that means that 10% of the time it will be wrong and say you do have it).

We want to find  $P(A|B)$ . Using the definition of conditional probability and the formula for the probability of two events occurring, we can write:

$$P(A \cap B) = P(A) \cdot P(B|A) \quad (1)$$

$$\frac{P(A \cap B)}{P(B)} = P(A|B) \quad (3)$$

Substituting 1 into 3, we obtain Bayes' Theorem:

$$P(B|A) \cdot \frac{P(A)}{P(B)} = P(A|B) \quad (7)$$

However, we still don't know  $P(B)$ . This can be easily computed with the total probability theorem:

$$P(B) = P(B|A) \cdot P(A) + P(B|\neg A) \cdot P(\neg A) = .99 \cdot .1 + .1 \cdot (1 - .1) = .189$$

This yields an answer of  $.99 \cdot .1 / .189 \approx .52$

You may be surprised by how low this number is compared to the accuracy of the test, which is  $.99 \cdot .1 + .9 \cdot .9 = .909$ . This means that just because a test has a high accuracy does not necessarily mean that you should panic if you test positive (but please stay home if you have the COVID-19).

Bayes' theorem is a simple formula, but it is very powerful since it relates an observed event (B) to the conditions that caused that event. This is the type of thinking you go through whenever you do a science experiment: you run tests and record data, and then use that data to draw conclusions about the system you tested. This is the backbone of the theory of Bayesian Statistics, which has many applications in signal processing, science research, game theory, and more. (In the problem set, look out for problems labeled "inference," as these problems relate to Bayesian statistics.)

## 5 Problems

Now that you are an expert on conditional probability, try out these problems!

1. Mario has two children. Assume that children are equally likely to be born as a boy or a girl and are equally likely to be born on any day of the week. I ask Mario if he has a daughter, and he says yes. What is the probability the other child is a son? What if instead I ask Jerry, who also has two kids, if he has a daughter that was born on a Tuesday, and he says yes; what is the probability the other child is a son in this scenario?
2. **Inference #1:** Reimu has 2019 coins  $C_0, C_1, \dots, C_{2018}$ , one of which is fake, though they look identical to each other (so each of them is equally likely to be fake). She has a machine that takes any two coins and picks one that is not fake. If both coins are not fake, the machine picks one uniformly at random. For each  $i = 1, 2, \dots, 1009$ , she puts  $C_0$  and  $C_i$  into the machine once, and the machine picks  $C_i$ . What is the probability that  $C_0$  is fake? (HMMT Feb 2019 Guts #13)
3. **Prediction #1:** A bag contains nine blue marbles, ten ugly marbles, and one special marble. Ryan picks marbles randomly from this bag with replacement until he draws the special marble. He notices that none of the marbles he drew were ugly. Given this information, what is the expected value of the number of total marbles he drew? (HMMT Feb 2018 Combo #5)
4. **Prediction #2:** Noted magician Casimir the Conjurer has an infinite chest full of weighted coins. For each  $p \in [0, 1]$ , there is exactly one coin with probability  $p$  of turning up heads. Kapil the Kingly draws a coin at random from Casimir the Conjurer's chest, and flips it 10 times. To the amazement of both, the coin lands heads up each time! On his next flip, if the expected probability that Kapil the Kingly flips a head is written in simplest form as  $\frac{p}{q}$ , then compute  $p + q$ . (PUMaC 2018 Live Round Calculus #1)
5. **Inference #2:** Johnny has a deck of 100 cards, all of which are initially black. An integer  $n$  is picked at random between 0 and 100, inclusive, and Johnny paints  $n$  of the 100 cards red. Johnny shuffles the cards and starts drawing them from the deck. What is the least number of red cards Johnny has to draw before the probability that all the remaining cards are red is greater than .5?
6. **Inference #3:** Yannick picks a number  $N$  randomly from the set of positive integers such that the probability that  $n$  is selected is  $2^{-n}$  for each positive integer  $n$ . He then puts  $N$  identical slips of paper numbered 1 through  $N$  into a hat and gives the hat to Annie. Annie does not know the value of  $N$ , but she draws one of the slips uniformly at random and discovers that it is the number 2. What is the expected value of  $N$  given Annie's information? (HMMT Feb 2019 Guts #29) (Note: Uses Calculus)

## Citations

Foundation, CK-12. "12 Foundation." CK, [www.ck12.org/book/cbse-maths-book-class-12/section/14.5/](http://www.ck12.org/book/cbse-maths-book-class-12/section/14.5/).