Advanced Statistics - Conditional Probability

Ralf Becker

January 23, 2023

1 Warming-up - Monty Hall problem

You are on a game show, being asked to choose between three doors. Behind one door is a car and behind the others are goats. You choose a door, say Door 1. The host, Monty Hall, picks one of the other doors, which he knows has a goat behind it, and opens it, showing you the goat. (You know, by the rules of the game, that Monty will always reveal a goat.) Monty then asks whether you would like to switch your choice of door to the other remaining door.
The big question: assuming you prefer having a car more than having a goat, do you choose to switch or not to switch?
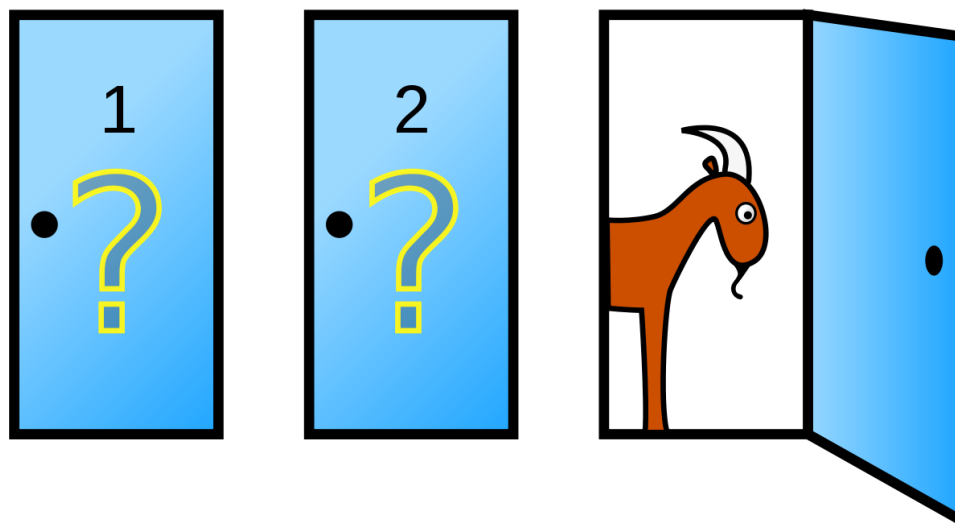

Figure 1: Monty Hall problem

The solution is that switching will let you win **twice** as often as sticking with the original choice, a result that seems counter-intuitive to many. In the following, we will introduce conditional probability and Bayes' Theorem, which give us a way to explain this result.

2   Conditional Probability

An important consideration in the development of probability is that of *conditional probability.* This refers to the calculation of updating probabilities in the light of revealed information. For example, insurance companies nearly always set their home contents insurance premiums on the basis of the postcode in which the home is located. That is to say, insurance companies believe the risk depends upon the location; i.e., the probability of property crime is assessed conditional upon the location of the property. (A similar calculation is made to set car insurance premiums.) As a result, the premiums for two identical households located in different parts of the country can differ substantially.

- In general, the probability of an event, E, occurring *given* that an event, F, has occurred is called the *conditional probability* of E given F and is denoted $\Pr(E|F)$.

Let's consider all young people of the same age, say 23 years old. That will be approximately your age after graduating from University. Apart from the intellectual challenge and pleasure you get from studying. Many of you would be studying in order to improve your job market prospect. Let us define the event $F = $ *a person has a University degree.* Then we could think of different types of events to describe job market success. Say that you have a job, or that you have annual earnings above £30,000 or that you have a job that excites you. For now, let's stick with the event definition $E = $ *a person is in employment.* You as a student are certainly hoping that $\Pr(E|F) > \Pr(E|\bar{F})$.

As a preliminary to the main development, consider the simple experiment of rolling a fair die and observing the number of dots on the upturned face. Then $S = \{1, 2, 3, 4, 5, 6\}$ and define events, $E = \{4\}$ and $F = \{4, 5, 6\}$; we are interested in $\Pr(E|F)$. To work this out we take F as known. Given this knowledge the sample space becomes restricted to simply $\{4, 5, 6\}$ and, given no other information, each of these three outcome remains equally likely. So the required event, 4, is just one of three equally likely outcomes. It therefore seems reasonable that $\Pr(E|F) = \frac{1}{3}$.
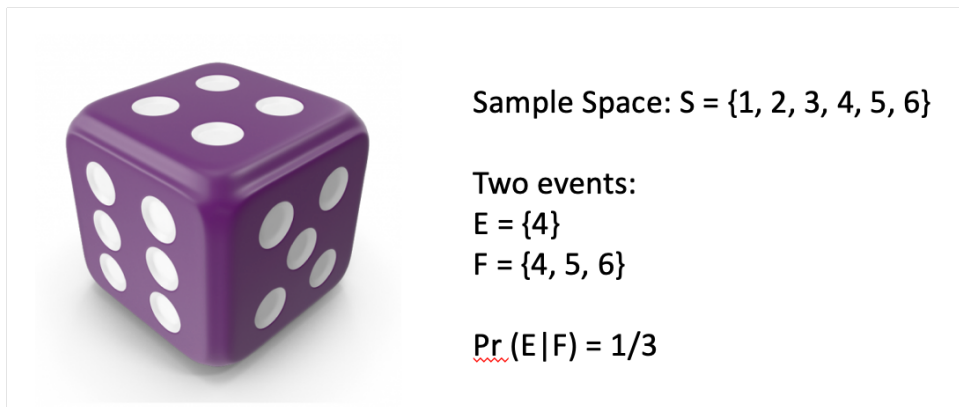
Figure 2: Conditional probability - dice example

We shall now develop this idea more fully, using Venn Diagrams with the implied notion of area giving probability. Consider an abstract sample space, denoted by S, with events $E \subset S$ and $F \subset S$. This is illustrated in the following Figure. Eventually we will want to construct the conditional probability, $\Pr(E|F)$. Sticking with the above example that could be the probability that *you are employed at the age of 23*, given that *you have a university degree*. Two important areas used in the construction of this conditional probability are highlighted as **a** and **b**:
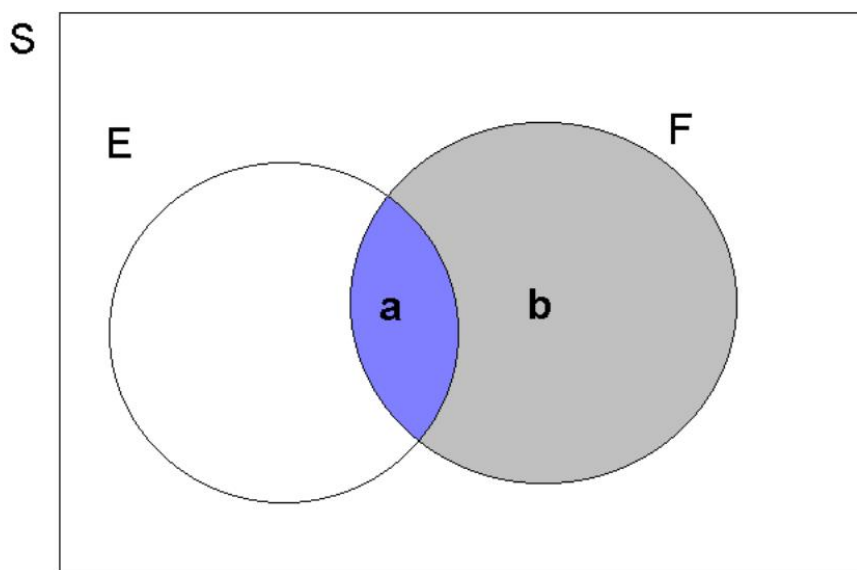

Figure 3: Conditional proability

In general, it is useful to think of $\Pr(E)$ as $\frac{area(E)}{area(S)}$; and similarly for $\Pr(F)$. The $\Pr(E \cap F)$ could equally be thought of as $\frac{area(a)}{area(S)}$. With this in mind, consider what happens if we are

now told that F has occurred. Incorporating this information implies that the effective sample space becomes restricted to $S^* = F$, since F now defines what can happen. This now covers the sample area $a + b$. On this new, restricted, sample space an outcome in E can only be observed if that outcome also belongs to F, the restricted sample space S^*. And this only occurs in area a which corresponds to the event $E \cap F$. Thus the event of interest *now* is $E^* = E \cap F$, as defined on the *restricted* sample space of $S^* = F$.

In order to proceed with the construction of the conditional probability, $\Pr(E|F)$, let $area(S) = z$. Then, since the ratio of the area of the event of interest to that of the sample space gives probability, we have (on this restricted sample space):

$$
\begin{aligned}
\Pr(E|F) &= \frac{area\,(E \cap F)}{area\,(F)} \\
&= \frac{a}{a+b} \\
&= \frac{a/z}{(a+b)/z} \\
&= \frac{\Pr\,(E \cap F)}{\Pr\,(F)},
\end{aligned}
$$

We have shown, for this example how a conditional probability can be expressed as a function of the joint probability $\Pr(E \cap F)$ and the probability $\Pr(F)$. We call the probability $\Pr(F)$, which is neither a joint probability nor a conditional probability, a *marginal* probability. This is a profound result and should be formulated in more general terms:

Definition. The probability that E occurs, given that F is known to have occurred, gives the **conditional probability** of E given F. This is denoted $Pr(E|F)$ and is calculated as

$$
\Pr(E|F) = \frac{\Pr(E \cap F)}{\Pr(F)}
$$

and from the axioms of probability will generate a number lying between 0 and 1, since $\Pr(F) \geq \Pr(E \cap F) \geq 0$.

Example. *A Manufacturer of electrical components knows that the probability is 0.8 that an order will be ready for shipment on time and it is 0.6 that it will also be delivered on time. What is the probability that such an order will be delivered on time given that it was ready for shipment on time?*

Let $R = READY$, $D = DELIVERED\ ON\ TIME$. $Pr(R) = 0.8$, $Pr(R \cap D) = 0.6$. From this we need to calculate $Pr(D|R)$, using the above formula. This gives, $Pr(D|R) = Pr(R \cap D)/Pr(R) = 6/8$, or, 75%.

If we re-arrange the above formula for conditional probability, we obtain the so-called *multiplication rule of probability* for *intersections* of events:

2.1 Multiplication rule of probability

The multiplication rule of probability can be stated as follows:

$$\Pr(E \cap F) = \Pr(E|F) \times \Pr(F) \tag{1}$$

It is equally true that

$$\Pr(E \cap F) = \Pr(F|E) \times \Pr(E) \tag{2}$$

Note that for any two events, E and F, $(E \cap F)$ and $(E \cap \bar{F})$ are mutually exclusive; they were areas a and b respectively in the above Venn diagram. Also, $E = (E \cap F) \cup (E \cap \bar{F})$; this has been seen before. So the *addition rule* and *multiplication rule* of probability together give:

$$\begin{aligned} \Pr(E) &= \Pr(E \cap F) + \Pr(E \cap \bar{F}) \\ &= \Pr(E|F) \times \Pr(F) + \Pr(E|\bar{F}) \times \Pr(\bar{F}). \end{aligned}$$

This is an extremely important and useful result, in practice, as we shall see shortly.

The following examples refer to a jar with different coloured marbles. So what you should be picturing is something like the following:
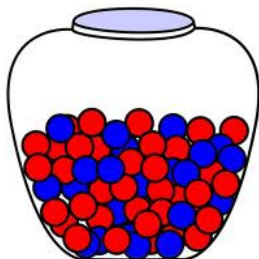
Example. *A jar contains 6 red marbles and 4 blue marbles. Two marbles are drawn from the bag, without replacement. What is the probability that both marbles are blue?*   2/15

Here are the steps:
Step 1: Label your events A and B. Let A be the event that marble 1 is blue and let B be the event that marble 2 is blue. You want to calculate $\Pr(B \cap A) = \Pr(B|A) \times \Pr(A)$
Step 2: Figure out the probability of A. There are ten marbles in the bag, so the probability of drawing a blue marble is $\Pr(A) = 4/10$.
Step 3: Figure out the probability of B given the first marble was blue, $\Pr(B|A)$. There are nine marbles left in the bag, and if the first marble was blue then there are only three left which are blue. So the probability of choosing a blue marble $Pr(B|A)$ is 3/9.

Step 4: Multiply Step 2 and 3 together: $(4/10) * (3/9) = 2/15$.

Example. *A jar contains 4 red marbles, 4 green marbles, and 5 blue marbles. If we choose a marble, then another marble without putting the first one back in the jar, what is the probability that the first marble will be blue and the second will be green?*   5/39

Hint:

Step 1: The probability of event A happening, then event B, is the probability of event A happening times the probability of event B happening given that event A already happened, $\Pr(B \cap A) = \Pr(B|A) \times \Pr(A)$. *In this case, event A is picking a blue marble and leaving it out. Event B is picking a green marble in the second draw.*

Step 2: Let's take the events one at at time. What is the probability that the first marble chosen will be blue?

Step 3: There are 5 blue marbles, and 13 total, so the probability we will pick a blue marble is $\Pr(A) = 5/13$.

Step 4: After we take out the first marble, we don't put it back in, so there are only 12 marbles left.

Step 5: Since the first marble was blue, there are still 4 green marbles left.

Step 6: So, the probability of picking a green marble after taking out a blue marble is $Pr(B|A) = 4/12$.

Step 7: Therefore, the probability of picking a blue marble, then a green marble is $(5/13)(4/12) = 5/39$.

2.1.1   Additional resources

Khan Academy

- Another application of this rule [https://www.khanacademy.org/math/probability/independent-dependent-probability/dependent_probability/v/dependent-probability-example-1]

2.2   Statistical Independence

If the knowledge that F has occurred does NOT alter our probability assessment of E, then E and F are said to be (statistically) *independent*. In this sense, F carries no information about E.

Definition. Formally, E and F are **independent** events if and only if

$$Pr(E|F) = Pr(E)$$

which, in turn is true *if and only if*

$$Pr(E \cap F) = Pr(E) \times Pr(F).$$

As you can see, the multiplication rule of probabilities, in the case of independent events, is $Pr(E \cap F) = Pr(E) \times Pr(F)$ whereas in general it is $Pr(E \cap F) = Pr(E) \times Pr(F|E)$.
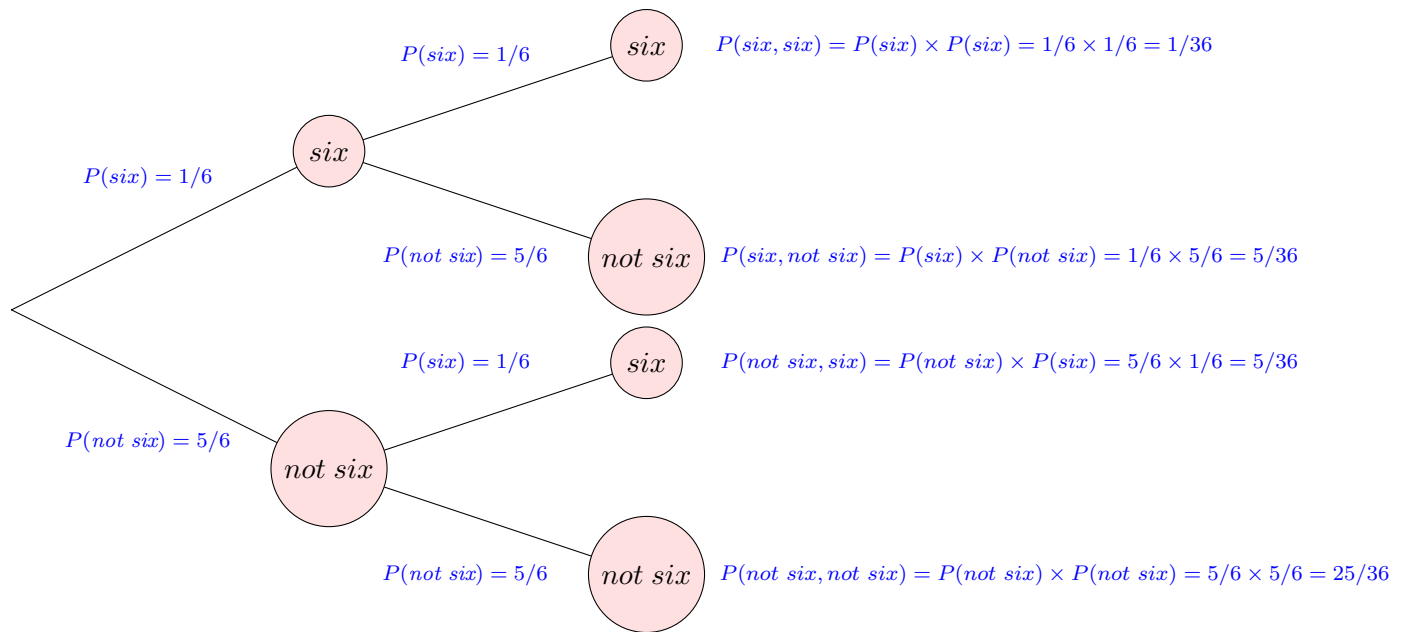
This concept of independence is of enormous importance in practice. Consider the case of lung cancer and its connection to smoking (apologies to all smokers for being picked upon here). The first connection between smoking and lung cancer was made in the 1920s. However, for many decades, with the tobacco industry spending a lot of money and effort to convince people that there was no connection between the two, there was no certainty about a causal connection. In other words, the tobacco industry claimed that the two events are *independent*, or $Pr(Cancer|Smoking) = Pr(Cancer|Smo\bar{o}king) = Pr(Cancer)$. It was then the task of epidemiologists to show otherwise. This was famously and comprehensively achieved by the [`http://en.wikipedia.org/wiki/British_Doctors_Study` *British Doctors Study*].

In the smoking - lung cancer example the question is really whether smoking **causes** lung cancer. It has now been established that this is indeed the case. That however, is not the same as the two events being dependent. There can be statistical dependence between two events without there being such a causal relationship. For instance, the number of umbrellas being sold in a year could well be positively related to the size of the grain harvest in a country. So, statistically these two variables are likely to be dependent. But that does not imply that people buying more umbrellas increases the grain harvest. The reason the two are related to each other is that they are both dependent on a common third variable, the weather. Questions of causality are super important in economics and the sub-field of econometrics thinks very carefully about methods that can be used to establish causal relationships (which are a sub-group of correlation relationships). In other words: not all correlations do correspond to causal relationships.

Example. *Two fair dice are rolled. Use a probability tree diagram to determine the probability of obtaining:*
(a) two sixes,
(b) no sixes,
(c) exactly one six.

Before you start calculating, think about whether the first and the second roll of the dice are independent.

$$P(six) = 1/6 \qquad six \qquad P(six, six) = P(six) \times P(six) = 1/6 \times 1/6 = 1/36$$

$$six$$

$$P(six) = 1/6$$

$$P(not\ six) = 5/6 \qquad not\ six \qquad P(six, not\ six) = P(six) \times P(not\ six) = 1/6 \times 5/6 = 5/36$$

$$P(six) = 1/6 \qquad six \qquad P(not\ six, six) = P(not\ six) \times P(six) = 5/6 \times 1/6 = 5/36$$

$$P(not\ six) = 5/6$$

$$not\ six$$

$$P(not\ six) = 5/6 \qquad not\ six \qquad P(not\ six, not\ six) = P(not\ six) \times P(not\ six) = 5/6 \times 5/6 = 25/36$$

(a) two sixes: $P(six, six) = 1/36$

(b) no sixes: $P(not\ six, not\ six) = 25/36$

(c) exactly one six: $P(six, not\ six) + P(not\ six, six) = 5/36 + 5/36 = 10/36$

You would have possibly solved the above example without having been aware of the concept of independence. But once you understand independence we can reformulate the above. Let $D1_6$ be the event for the first dice roll showing a 6 and $D2_6$ for the second dice roll showing a 6. After recognising that the outcome of the second dice roll will be independent of the outcome of the first dice roll, the probability for two sixes can now be calculated using the multiplication rule specialised for independent events:

$$P(D1_6 \cap D2_6) = P(D1_6) \times P(D2_6) = (1/6) \times (1/6) = 1/36$$

Equally, you can calculate the probability to not have any sixes:

$$P(D1_{1-5} \cap D2_{1-5}) = P(D1_{1-5}) \times P(D2_{1-5}) = (5/6) \times (5/6) = 25/36$$

For the probability to get exactly one six we will, as in the example above have to combine the two ways in which this can happen ($D1_6$ and then $D2_{1-5}$ or $D1_{1-5}$ and then $D2_6$).

$$
\begin{aligned}
P(D1_6 \cap D2_{1-5}) + P(D1_{1-5} \cap D2_6) &= P(D1_6) \times P(D2_{1-5}) + P(D1_{1-5}) \times P(D2_6) \\
&= (1/6)(5/6) + (5/6)(1/6) = 10/36
\end{aligned}
$$

Example. *On a Roulette Table you have numbers 0 to 36. Of the numbers 1 to 36 18 numbers are red and another 18 are black. The number 0 is white. You can assume that consecutive spins (S) of the roulette wheel produce independent outcomes. Calculate the probabilities for the following events:*

   (a) One red and one black number in two spins, the order not being important.

   (b) Three black numbers in a row.

   (c) The first spin producing a 0, followed by a red and then a black number.

   (d) The first spin producing a 0, followed by one red and black where the order is not important.

SOLUTION:

   (a) $P(rb) = (P(S1_b) \times P(S2_r)) + (P(S1_r) \times P(S2_b)) = (18/37)(18/37) + (18/37)(18/37) = 0.4733$

   (b) $P(S1_b \cap S2_b \cap S3_b) = P(S1_b) \times P(S2\_b) \times P(S3\_b) = (18/37)(18/37)(18/37) = 0.1151$

   (c) $P(S1_0 \cap S2_r \cap S3_r) = P(S1_0) \times P(S2\_r) \times P(S3\_r) = (1/37)(18/37)(18/37) = 0.0064$

   (d) $P(S1_0 \cap rb) = P(S1_0) \times P(rb) = (1/37) \times 0.4733 = 0.0128$

2.3   Bayes' Theorem

One area where conditional probability is extremely important is that of clinical trials - testing the power of a diagnostic test to detect the presence of a particular disease or infection. Suppose, then, that a new test is being developed and let $P =$ '*test positive*' and $D =$ '*presence of disease*', but where the results from applying the diagnostic test can never be wholly reliable. From the point of view of our previous discussion on conditional probability, we would of course require that $\Pr(P|D)$ to be large; i.e., the test should be effective at detecting the disease. However, if you think about, this is not necessarily the probability that we might be interested in from a diagnosis point of view. Rather, we should be more interested in $\Pr(D|P)$, the probability of correct diagnosis, and require this to be large (with, presumably, $\Pr(D|\bar{P})$ being small).

Note that in this section we are using $Pr(A)$ to denote the probability that event A occurs, where in previous sections we used $P(A)$. Both notations are commonly found and you need to be familiar with either. The reason why we switch on this occasion is as we are using an event we define as P and hence we use Pr to represent the probability to avoid confusion.

Here, what we are trying to attach a probability to is a possible 'cause'. The observed outcome is a positive test result (P), but the presence or non-presence of the disease is what is of interest and this is uncertain. $\Pr(D|P)$ asks the question '*what is the probability that it is the presence of the disease which caused the positive test result*'? Firstly, is $\Pr(D|P)$ high or low? Secondly, might there being something else (F) which could offer a "better" explanation, such that $\Pr(F|P) > \Pr(D|F)$?)

The situation is depicted in the following Figure, in which there are two possible 'states' in the population: D (depicted by the lightly shaded area covering the left portion of the sample space) and $\bar{D}$. It must be that $D \cup \bar{D} = S$, since any individual in the population either has the disease or does not. The event of an observed positive test result is denoted by the closed loop, P. (Notice that the shading in the diagram is relatively darker where P intersects with D.)
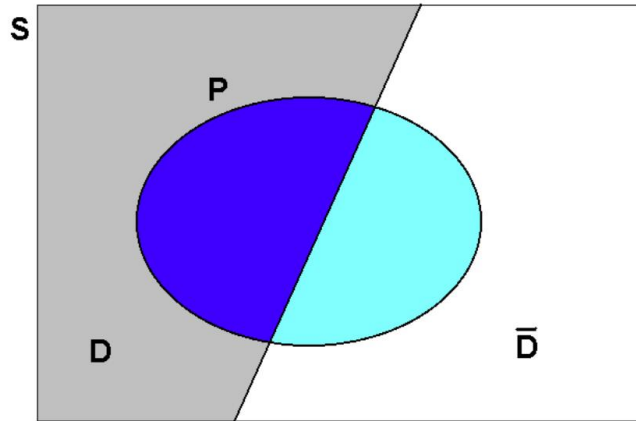

Figure 4: Bayes' Theorem

How might we construct the required probability, $\Pr(D|P)$? We start from the multiplication rule, $\Pr(D \cap P) = \Pr(D|P)\Pr(P)$, and rearrange it as follows:

$$
\begin{aligned}
\Pr(D|P) &= \frac{\Pr(D \cap P)}{\Pr(P)} \\
&= \frac{\Pr(D \cap P)}{\Pr(P \cap D) + \Pr(P \cap \bar{D})},
\end{aligned}
$$

since $P(P) = (P \cap D) \cup (P \cap \bar{D})$, and these are mutually exclusive. From the multiplication rule of probability, $\Pr(P \cap D) = \Pr(P|D) \times \Pr(D)$, and similarly for $\Pr(P \cap \bar{D})$. Thus

$$
\Pr(D|P) = \frac{\Pr(P|D) \times \Pr(D)}{\Pr(P|D) \times \Pr(D) + \Pr(P|\bar{D}) \times \Pr(\bar{D})},
$$

which may be convenient to work with since $\Pr(P|D)$ and $\Pr(P|\bar{D})$ can be estimated from clinical trials and $\Pr(D)$ estimated from recent historical survey data. We sometimes call probabilities like $Pr(D)$ **prior probabilities** and probabilities like $\Pr(D|P)$ **posterior probabilities**. You can think of the **prior** and **posterior** representing before and after the revealing of the information P.

This is all pretty abstract,so let's work through an example.

Example. *Let's assume that you are worried about having contracted Covid. You know that at the time around 10% of the population have Covid. This number comes from a regular survey undertaken by the Office for National Statistics. We also call this the prevalence of the disease, $P(D) = 0.1$. You want to take a test. You did your research about the test and found that, from clinical tests, it is known that the test has an accuracy of 95%. In this context this implies that, if you do have disease, then there is a 95% probability that the test will tell yo that you are positive. Also, if you do not have the disease the test will also tell you, with probability of 95% that you are not positive (or in 5% of cases it would tell a non-infected person that they are positive). These two probabilities do not need to be identical (see example below). In statistical language we know that $\Pr(P|D) = 0.95$ and $\Pr(P|\bar{D}) = 1 - 0.95 = 0.05$.*

After doing all this research you actually get to take the test. The test tells you that you are positive. Given all this information, what is the probability that you do have Covid? 100%? 95%? Neither of these, the probability is significantly lower. There are two ways how you could have ended up testing positive. Either you have Covid and, correctly tested positive, or you do not have Covid and the test erroneously showed a positive result. So, given you did test positive, what is actually the probability that you have Covid, $Pr(D|P)$. Just to benchmark this. If you just had not tested, the probability would be 10%, the population prevalence.

This is where the previous result comes in useful, as we actually do have all the terms required:

$$
\begin{aligned}
\Pr(D|P) &= \frac{\Pr(P|D) \times \Pr(D)}{\Pr(P|D) \times \Pr(D) + \Pr(P|\bar{D}) \times \Pr(\bar{D})} \\
&= \frac{0.95 \times 0.1}{0.95 \times 0.1 + 0.05 \times 0.9} = 0.6786
\end{aligned}
$$

So, this is certainly much less than certainty. But you should also note that after the positive test the probability that you have Covid increased substantially from 10% (before the test) to almost 68%.

Example. *The context if this exercise is the same as in the above example. But now the probabilities are:*

- *The disease prevalence is 5%.*

- *If you have the disease, the probability of the test showing as positive is 90%.*

- *If you do not have the disease, the probability of the test showing as positive is 15%.*

What are $Pr(D|P)$ and $Pr(D|\bar{P})$?

SOLUTION:

$$
\begin{aligned}
\Pr(D|P) &= \frac{\Pr(P|D) \times \Pr(D)}{\Pr(P|D) \times \Pr(D) + \Pr(P|\bar{D}) \times \Pr(\bar{D})} \\
&= \frac{0.90 \times 0.05}{0.90 \times 0.05 + 0.15 \times 0.95} = 0.2400.
\end{aligned}
$$

$$
\begin{aligned}
\Pr(D|\bar{P}) &= \frac{\Pr(\bar{P}|D) \times \Pr(D)}{\Pr(\bar{P}|D) \times \Pr(D) + \Pr(\bar{P}|\bar{D}) \times \Pr(\bar{D})} \\
&= \frac{0.10 \times 0.05}{0.10 \times 0.05 + 0.85 \times 0.95} = 0.0062.
\end{aligned}
$$

The two probabilities you calculated in this last example are very useful to understand how the test result updates your information. Before you test you start out with the the population prevalence of 5% as the probability that you have the disease. Once you receive the test result that probability increases if the test result is positive, $P(D|P) > P(D)$, or decreases if the test result is negative, $P(D|\bar{P}) < P(D)$. But such a test does not provide certainty.

This sort of calculation is an example of *Bayes' Theorem.* Of course, we may have to consider more than two possible causes, and the construction of the appropriate probabilities is as follows.

1. Consider a sample space, S, where $E \subset S$ and A, B, C are three mutually exclusive events (possible causes), defined on S, such that $S = A \cup B \cup C$. In such a situation, A, B and C are said to form a **partition** of S.
   Bayes' Theorem states that:

   $$
   \Pr(A|E) = \frac{\Pr(E|A) \times \Pr(A)}{\{\Pr(E|A) \times \Pr(A)\} + \{\Pr(E|B) \times \Pr(B)\} + \{\Pr(E|C) \times \Pr(C)\}}.
   $$

2. And, more generally, consider a sample space, S, where $E \subset S$ and $F_1, F_2, ..., F_k$ are k mutually exclusive events (possible causes), which form a partition of $S : S = \bigcup_{j=1}^{k} F_j$.
   Bayes' Theorem then states that:

   $$
   \Pr(F_j|E) = \frac{\Pr(E|F_j) \times \Pr(F_j)}{\sum_{s=1}^{k} \{\Pr(E|F_s) \times \Pr(F_s)\}}.
   $$

From the above formula, you should be able to satisfy yourself that $\sum_{j=1}^{k} \Pr(F_j|E) = 1$. If this is not at first clear, consider case (1) and show that $\Pr(A|E) + \Pr(B|E) + \Pr(C|E) = 1$. The reason for this is that since A, B and C form a partition of S, they must also form a partition of any event $E \subset S$. In the above conditional probabilities, we are regarding E as the restricted sample space and therefore the probabilities assigned the mutually exclusive events (A, B, C) which *cover* this (restricted) sample space, E, must sum to 1.

Example. : *Box A contains 2 red balls. Box B contains 1 red and 1 white ball. Box A and Box B are identical. If a box is selected at random and one ball is withdrawn from it, what is the probability that the selected box was A if the ball withdrawn from it turns out to be red?*

Let A be the event of selecting Box A and R the event of drawing a red ball. Require $Pr(A|R)$.
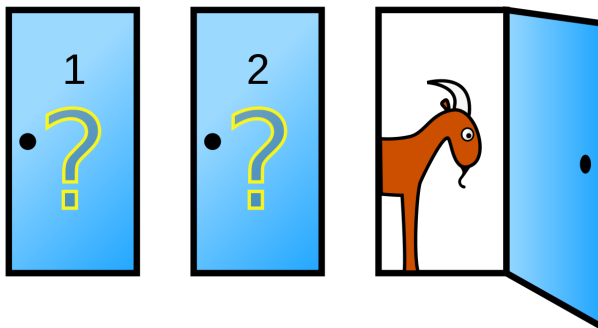
$Pr(A|R) = Pr(A \cap R)/Pr(R);$

$Pr(A \cap R) = Pr(A)Pr(R|A) = (1/2) \times 1 = 1/2.$

And,

$$
\begin{aligned}
Pr(R) &= Pr(A \cap R) + Pr(\bar{A} \cap R) \\
&= \Pr(A) \times \Pr(R|A) \quad + \quad \Pr(\bar{A}) \times \Pr(R|\bar{A}) \\
&= (1/2) \quad + \quad (1/2) \times (1/2) \\
&= 3/4.
\end{aligned}
$$

Therefore, $\Pr(A|R) = (1/2)/(3/4) = 2/3$.

Now we come back to the Monty Hall problem which has been introduced at the beginning.

Recall where we ended up in the show. You had selected one door, say Door 1, hoping that you would win a car if that was hidden behind that door. The show's host opened one of the other doors revealing a goat (here Door 3). Being an avid viewer of the show you know that the host,

at that stage would never actually open the door with the car. Should you switch from Door 1 to 2, or does it not matter for your chances of winning a car?

As indicated at the beginning, switching will let you win twice as often as sticking with the original choice. We now have the statistical tools to investigate why this is the case. We could use a table to explicitly list out all the possible outcomes, and count how often you get the car if you stay versus switch. Without loss of generality, suppose your selection was door 1. Then the possible outcomes can be seen in this table:

Behind door 1	Behind door 2	Behind door 3	Result if staying at door 1	Result if switching to the door offered
Car	Goat	Goat	**Wins car**	Wins goat
Goat	**Car**	Goat	Wins goat	**Wins car**
Goat	Goat	**Car**	Wins goat	**Wins car**

In two out of three cases, you win the car by changing your selection after one of the doors is revealed. This is because there is a greater probability that you choose a door with a goat behind it in the first go, and then Monty is guaranteed to reveal that one of the other doors has a goat behind it. Hence, by changing your option, you double your probability of winning.

It is, admittedly, not straightforward to follow this logic and therefore we will appreciate the formal argument using **Bayes' theorem**. The version you see below ids only one of several other ways in which you could see the problem through the Bayes' Theorem lens. But all different perspectives will lead you to the same outcome. We argued earlier that Bayes' theorem allows us to think how new information allows us to update probabilities using the new evidence. When you chose Door 1 initially you knew that the probability of the car being behind Door 1 was 1/3, $Pr(D1) = 1/3$, and also $Pr(D2) = Pr(D3) = 1/3$.

Now we describe how you can update the probability $Pr(D1)$ with the new information obtained after the quizmaster opened Door 3, event $= OD3$. The following assumes that you have chosen Door 1. Equivalent considerations would apply if you had chosen doors 2 or 3. The Bayes formula adjusted to the new event names looks as follows:

$$\Pr(D1|OD3) = \frac{\Pr(OD3|D1) \times \Pr(D1)}{\Pr(OD3|D1) \times \Pr(D1) + \Pr(OD3|D2) \times \Pr(D2) + \Pr(OD3|D3) \times \Pr(D3)}$$

We already know $Pr(D1) = Pr(D2) = Pr(D3) = 1/3$. Further we need:

- $\Pr(OD3|D1)$ is the probability that Monty shows Door 3, given $(D1)$, i.e. that the car is behind Door 1. That probability would 1/2 as Monty has the choice between Doors 2 and 3 as both hide a goat.

- $\Pr(OD3|D2)$ is the probability that Monty shows Door 3, given that there is a car behind Door 2. As you have chosen Door 1 Monty cannot open Door 1 and he also cannot open

14

Door 2 as that is where the car is. Therefore he is certain to open Door 3, $Pr(OD3|D2) = 1$.

- $\Pr(OD3|D3)$ is the probability that Monty shows Door 3, given that the car is behind Door 3. Monty would never do this, as he does not want the game to end at this stage. Therefore $Pr(OD3|D3) = 0$.

Combining all of this information gives

$$\Pr(D1|OD3) = \frac{(1/2) \times (1/3)}{(1/2) \times (1/3) + 1 \times (1/3) + 0 \times (1/3)} = 1/3$$

So the probability that the car is behind Door 1 has not changed at all through Monty's act of revealing a goat behind Door 3. At this stage you know that the car is not behind Door 3 and therefore the only other possibility is that the car is behind Door 2. Logically the probability, at this stage, that the car is behind Door 2 ought to be $Pr(D2|OD3) = 2/3$. This is what can be somewhat unintuitive, but let's update the initial $Pr(D2) = 1/3$ with the information provided by the opening of Door 3, again using Bayes's Theorem:

$$\begin{aligned}\Pr(D2|OD3) &= \frac{\Pr(OD3|D2) \times \Pr(D2)}{\Pr(OD3|D1) \times \Pr(D1) + \Pr(OD3|D2) \times \Pr(D2) + \Pr(OD3|D3) \times \Pr(D3)} \\ &= \frac{1 \times (1/3)}{(1/2) \times (1/3) + 1 \times (1/3) + 0 \times (1/3)} = 2/3\end{aligned}$$

Opening Door 3 has doubled the probability that the car is behind Door 2!!! Therefore, switching to Door 2 is twice as likely to get you the car as staying with Door 1. You can repeat this argument for any other initial choice of Door. You will always find that switching increases the probability of you winning the car.

In this Section you learned that conditional probabilities are important to describe important relationships in the society, such as the probabilities of contracting Covid conditional on being vaccinated, the probabilities of cars being stolen depending on where they are parked and the probabilities of having an interesting job, depending on what type of education you have. In addition you learned that, using the incredibly important Bayes' Theorem, you can solve rather complicated probabilistic problems.

3 Additional resources

Khan Academy

- A different example that intuitively leads to Bayes Formula [`https://www.khanacademy.org/math/probability/independent-dependent-probability/dependent_probability/v/introduction-to-dependent-probability`]

- Should you switch the door? A classic problem. [`https://www.khanacademy.org/math/probability/independent-dependent-probability/dependent_probability/v/monty-hall-problem`]

4 Footnotes