

## 2 Probability

### 2.1 Why probability?

At a superficial level, it would appear that the study of probability is just the reverse of statistics. In the former, we know what the population is and the question is about a sample we take out of the population. We are trying to find out what the chances are that it is going to be of a particular kind (think of the standard example of a jar with different coloured marbles, and then having to calculate the probability of marbles drawn from this jar being of a particular colour). In statistics, we study the sample and then try to infer something about the population.

In fact the two are very closely related. As we shall see, the inferences that statistics allows us to make about the population will also be in the form of probabilities. The role of probability in economics goes well beyond its use in statistics. When there is a lack of perfect information, we resort to probabilistic ways of thinking about economic problems. The areas of risk and insurance, which form the basis of much of the financial industry, require a deep understanding of probability. We shall see a relevant example of this during the course of this lecture.

### 2.2 Some definitions

**Random trial or experiment:** A process whose outcome is not known with certainty.

The most common example of this is a coin toss with two possible outcomes  $H$  or  $T$ . We do not know which of these is going to happen for any given coin toss. There could be processes for which we not only do not know which outcome is going to happen, but we also do not know all the possible outcomes. We will

restrict ourselves to the first kind for now where we know all the possible outcomes but not which one of them is going to happen.

**Sample space:** The collection or set of all possible outcomes of a random trial.

Note that *sample space* is different from the population/sample nomenclature used in statistics. The elements of the sample space are the outcomes of the random trial. In the example of the coin toss the sample space  $S$  is  $\{H, T\}$ . There can be sample spaces where the outcomes vary on more than one dimension. For example, consider a cricket match being played between teams Blue and Green on a day with the chance of rain. The two variables are the result of the match and whether it rains or not. Hence, the set of all outcomes will be  $\{It\ rains\ and\ Blue\ wins, It\ rains\ and\ Green\ wins, It\ rains\ and\ the\ match\ is\ abandoned, It\ doesn't\ rain\ and\ Blue\ wins, It\ doesn't\ rain\ and\ Green\ wins, It\ doesn't\ rain\ and\ the\ match\ is\ abandoned\}$ .

**Event:** Any set of outcomes that are part of the sample space.

In the cricket example, an event  $A$  could be that it rains. Hence,  $A = \{It\ rains\ and\ Blue\ wins, It\ rains\ and\ Green\ wins, It\ rains\ and\ the\ match\ is\ abandoned\}$ . Another event  $B$  could be that Blue wins. Hence,  $B = \{It\ rains\ and\ Blue\ wins, It\ doesn't\ rain\ and\ Blue\ wins\}$ .

**$C=A$  and  $B$ :** The event  $C$  contains the outcomes that occur *both* in  $A$  and  $B$ .

In the terminology of set theory we would refer to  $C$  as the *intersection* of  $A$  and  $B$ , or  $C = A \cap B$ . Going by the definition of events  $A$  and  $B$  given above,  $C = \{It\ rains\ and\ Blue\ wins\}$ .

**$C=A$  or  $B$ :** The event  $C$  contains the outcomes that occur either in  $A$  or in  $B$  or in both.

In set theory we would refer to  $C$  as the *union* of  $A$  and  $B$ , or  $C = A \cup B$ .

Again, using the definition of events  $A$  and  $B$  given earlier,  $C = \{It\ rains\ and\ Blue\ wins,\ It\ rains\ and\ Green\ wins,\ It\ rains\ and\ the\ match\ is\ abandoned,\ It\ doesn't\ rain\ and\ Blue\ wins\}$ .

## 2.3 Probability

The probability of an event is a number assigned to that event that indicates the likelihood of that event happening i.e. the likelihood of one of the outcomes that form the event occurring. We denote the probability of an event  $A$  taking place as  $P(A)$ . There are three basic axioms of probability.

1. The probability of an event will always be a positive fraction lying between 0 and 1, with  $P(A) = 0$  indicating that the event will not occur with certainty and  $P(A) = 1$  indicating that the event will occur with certainty. Increasing values of  $P(A)$  indicate increasing likelihood of the event taking place. We can write this axiom in symbols as  $0 \leq P(A) \leq 1$ .
2. One of the outcomes in the sample space will always happen. This is true by definition as the sample space is the set of all possible outcomes. Hence,  $P(S) = 1$ .
3. If  $A$  is an event consisting of  $n$  outcomes, then let  $A_i, i = 1, 2, \dots, n$ , be events such that each includes one and only one of these outcomes. For example, when  $A$  defined as earlier, we would have  $A_1 = \{It\ rains\ and\ Blue\ wins\}$ ,  $A_2 = \{It\ rains\ and\ Green\ wins\}$  and  $A_3 = \{It\ rains\ and\ the\ match\ is\ abandoned\}$ . The probability of  $A$  happening will be the sum of the probabilities of  $A_1, A_2, \dots$  and  $A_n$ . Hence,  $P(A) = \sum_{i=1}^n P(A_i)$ .

How do we find the probability of an event? In general it is not straightforward, but in the special case when all outcomes are equally likely, the probability of an event is the number of outcomes in that event divided by the total number of

possible outcomes. For example, if we roll a fair six-sided dice, the sample space of the number turning up is  $\{1, 2, 3, 4, 5, 6\}$ . If event  $E$  is getting an even number, i.e.  $E = \{2, 4, 6\}$ , then  $P(E) = 3/6 = 0.5$ . (Can you derive this rule from the axioms above?)

But this is not applicable in general. For example, in the case of the cricket match, the probability of Blue winning is not simply 0.33 just because the event contains two of the six possible outcomes. The outcomes are not equally likely, hence just counting the outcomes doesn't work. In general we can assign probabilities to events if we know the underlying process that is generating the outcomes. For example, let us think of the following trial - we toss a fair coin and if we get a head, we toss the coin again. The sample space is  $\{T, HH, HT\}$ . The three outcomes are not equally likely, but because we know the process by which these outcomes were generated, we know that  $P(T) = 0.5$ ,  $P(HH) = 0.25$ , and  $P(HT) = 0.25$ .

If we think about two events together, then  $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$ . We need to subtract  $P(A \text{ and } B)$  from the sum of  $P(A)$  and  $P(B)$ , as outcomes that are common to both get counted twice. If we take the dice examples, we can define two events - event  $E$  is getting an even number, and  $L$  is getting a number larger than 3. We can see that while the probabilities of both  $E$  and  $L$  are 0.5 each, the probability of getting an even number or a number larger than three is approximately 0.67, as the outcomes 4 and 6 occur in both.

## 2.4 Probability and statistics

We intuitively understand that if on tossing a coin, the probability of getting heads is 0.5, then if we toss the coin a large number of times, roughly 50% of the coin tosses should result in heads. This intuition forms the basis of a group of statistical theorems called the Laws of large numbers. One form of the law says that if the same random trial is repeated a large number of times, then the fraction

of times an event occurs is close to the actual probability of that event. This leads to a statistical definition of probability. Probability of an event is the fraction of times the event occurs if the trial is repeated an infinite number of times. Let us now link it back to the statistical concept of sample and population. Think of the population as the infinite number of trials, and the sample as the finite number of trials that we actually observe. The law of large number tells us that if our sample is large enough, i.e. if we observe a large enough number of trials, then the fraction of times that an event occurs in our sample, is close to the fraction of times it occurs in the population. This is the basis of inferential statistics and the reason why statisticians are always fretting about sample size!

*Note: Sometimes the intuition that drives the law of large numbers leads us to make mistakes in ascribing probabilities. For example, if a coin is tossed four times, and comes up heads each time, we feel that the fifth time the probability of tails is higher as on average we must have an equal number of heads and tails. But this is incorrect. The probability of tails in any one coin toss is 0.5 by definition. Remember that it is a law of **large** numbers and is not applicable to any one trial.*

## 2.5 Conditional probability

Often we will need to think about the probability of an event  $B$ , when we know that another event  $A$  has already occurred or is sure to occur. This is called the probability of  $B$  conditional on  $A$  having occurred and is written as  $P(B|A)$ . Let us go back to the dice example, with the two events  $E$  and  $L$ . Suppose someone rolls the dice and tells you that the number is greater than 3, i.e.  $L$  has occurred. Now what is the probability of the number being even, i.e. what is  $P(E|L)$ . In this case it would be  $2/3$  or approximately 0.67, as out of the three possibilities, two are even. Another way to think about it is that when  $L$  has happened, i.e., the number is larger than 3, the new sample space is  $L$ , since the only possible outcomes are the outcomes that are in  $L$ . We know that the only way  $E$  can happen now is if the outcome is one that was common to both  $L$  and  $E$ , i.e. 4

and 6 in this case. Hence, the conditional probability of  $E$  when  $L$  has already happened is probability of the occurrence of one of the outcomes in  $L$  and  $E$  out of all the outcomes in  $L$ . Hence, we can write

$$P(E|L) = \frac{P(E \text{ and } L)}{P(L)}$$

In this case we can put values and check the result.  $P(L) = 1/2$  and  $P(E \text{ and } L) = 1/3$ , hence  $P(E|L) = 2/3$ .

Let us take go back to the cricket example. Event  $A$  is that it rains and event  $B$  is that Blue wins. Suppose the chance of rain in 20%, and when it rains, the chance of the Blue team winning is 60%. What is the probability that it rains and the Blue team wins? In this question, we have been told  $P(A) = 0.2$ , and  $P(B|A) = 0.6$ , and we have been asked to find  $P(A \text{ and } B)$ . The answer using the formula above is 0.12. Let us try to understand what is happening here. There is a 20% chance of rains. This means that with a probability of 0.2, one of three things will happen - it will rain and Blue will win, it will rain and Green will win, or it will rain and the match will be abandoned. Now, we are also told that within these three possibilities, the probability of the first happening, i.e. Blue winning, is 0.6. Hence, the overall probability of it raining *and* within the three 'rainy' possibilities, the first one happening is  $0.2 \times 0.6 = 0.12$ .

## 2.6 Independence

Suppose that in the earlier example, the chance of Blue winning, given that it *does not* rain is also 60%. So, if we denote not raining by  $A'$ , then  $P(B|A') = 0.6$ . Therefore, the probability of Blue team winning is not affected by whether it rains or not. Regardless of weather conditions, the probability of Blue winning is 0.6. Hence, the *unconditional* probability of Blue winning is 0.6, i.e.  $P(B) = 0.6$ . This implies that the probability of Blue team winning is *independent* of whether it rains or not.

Two events are said to be independent of each other if the conditional probability of one event with respect to the other is same as its unconditional probability, i.e.,  $A$  and  $B$  are independent if  $P(A|B) = P(A)$ . Using the formula of conditional probability, we get an important result.

$$P(A|B) = \frac{P(A \text{ and } B)}{P(B)}; P(A|B) = P(A)$$

$$\Rightarrow P(A \text{ and } b) = P(A)P(B)$$

Hence, the probability of two independent events occurring together is the product of the probabilities of the two events happening individually. Hence, if we toss a coin and roll a dice, then the probability of getting heads and an even number is  $0.5 \times 0.5 = 0.25$ . This is because the probability of getting heads is the same regardless of what number turns up on the dice and vice versa.

Now let us take a finance example. Suppose a bank gives out a loan and expects it to be paid back in one year. The amount due including interest is Rs 100,000 and the chance of default is 20%, i.e the bank expects to get back the due amount with probability 0.8. This would be considered a very risky loan! Suppose the bank makes another such loan. Hence the bank is due Rs 200,000 in a year's time. Now, the bank turns to an investor and says that it will offer the investor a great deal. The bank will combine the two loans and divide it into two *tranches*. The first Rs 100,000 received from either borrower will be in the first tranche and whatever is received later will be in the second tranche. The investor can buy the first tranche as a financial asset from the bank, i.e. when the first Rs 100,000 is paid, it will be given to the investor. The investor asks the bank, "What is the risk?". The bank says that only one of the borrowers needs to pay in for the investor to get the amount, the only way the investor doesn't get paid is if both borrowers default. The probability of that is  $0.2 \times 0.2 = 0.04$ . Hence, the investor will get paid with a probability of 96%! This is a much safer investment than the original loan. Is there a problem in this?

## 2.7 Bayes theorem

Suppose in the APU admission test, there are equal number of boys and girls. For boys the chances of an interview call are 10%. The overall percentage of applicants called for interview was 15%. What is the share of boys in the applicants interviewed?

There are two events here. One is being a boy in the applicant pool, let us call it event  $B$ , the probability of which is 0.5. The second is being called for interview, let us call it  $I$ , the probability of which is 0.15. We also know that if you are a boy, the probability of getting called is 10%, i.e.  $P(I|B) = 0.1$ . The question is given that one is called for an interview, what is the probability that it is a boy, i.e what is  $P(B|I)$ ? Let us do it by numbers first and then we will use the formula. Let the total number of applicants be 100. The number of boys is 50. The total number of applicants called for interview is 15. The total number of boys called for interview is  $50 \times 0.1 = 5$ . Hence, within the applicants called for interview, the share of boys is  $1/3$ , or 0.33.

We can now write down the formula for conditional probability.

$$P(B|I) = \frac{P(B \text{ and } I)}{P(I)}$$

We don't know what  $P(B \text{ and } I)$  is but we do know  $P(I|B)$ . So, we write down its formula.

$$P(I|B) = \frac{P(B \text{ and } I)}{P(B)}$$

Using these two, we can write

$$P(B \text{ and } I) = P(B|I)P(I) = P(I|B)P(B)$$

Hence,

$$P(B|I) = \frac{P(I|B)P(B)}{P(I)}$$

This is called Bayes theorem. If we substitute the appropriate values, we get  $0.1 \times 0.5 / 0.15 = 1/3$ .

Suppose there is a medical test for a disease. The test is 90% accurate, i.e. if a person with the disease takes the test, the test will give a 'positive' result 90% of the time. Similarly if a person without the disease takes the test, the test will give a 'negative' result 90% of the time. If you take the test and it turns out to be positive, does it mean that you have the disease with the probability 0.9? The answer is no! The question you are asking is very different from the information you have. If the event of a positive test result is  $T$ , and the event of having the disease is  $D$ , then what you know is that  $P(T|D) = 0.9$ . But the question you are asking is - what is  $P(D|T)$ ? These two can be very different as we saw in the earlier example. The example in the textbook (Floyd, Chapter 2) as well as in the video on Moodle explain this in detail.