Basics of Probability

WE'LL COVER THE FOLLOWING

For anyone taking their first steps in data science, probability is a must-know concept. In this lesson, we will learn this important piece of the puzzle by going through each concept in a simple way.

What Is Probability?

Probability is the numerical chance that something will happen; it tells us how likely it is that some event will occur.

Probability is one of those intuitive concepts that we use on a daily basis, without necessarily realizing that we are talking probability. Our lives are full of uncertainties; unless someone has superpowers to foresee the future, we don't know the outcomes of a particular situation or event until it actually happens. *Will I pass the exam with flying colors? Will it snow today? Will my favorite team win the match?* These are some examples of uncertain events. In statistical terms, "team won" is the *outcome* while "my team winning today's match" is the event. Probability is the measure of how likely this outcome is.

For example, if it is 80% likely that my team will win today, the probability of the outcome “the team won” for today’s match is 0.8; while the probability of the opposite outcome, “it lost”, is 0.2, i.e., *1 - 0.8*. Probability is represented as a number ***between 0 and 1***, where 0 indicates impossibility and 1 indicates certainty.

Why Is Probability Important?

With all the uncertainty and randomness that occurs in our daily life, probability helps us make sense of these uncertainties. It helps us understand the chances of various events. This, in turn, means that we can make informed decisions based on estimates or patterns of data collected previously. For example, if it is likely to rain, we can grab an umbrella before heading out. Or if a user is unlikely to check our app without a reminder, we can send them a notification.

How Does Probability Fit in Data Science?

Understanding the methods and models needed for data science, like logistic regression which we will encounter in the Machine Learning section, randomization in A/B testing, or experimental design, and sampling of data are examples of use-cases that require a good understanding of probability.

Calculating Probability of Events

Probability is a type of ratio where we compare how many times an outcome can occur compared to all possible outcomes. Simply put:

$$ProbabilityOfAnEventHappening = \frac{NumberOfWaysItCanHappen}{TotalNumberOfOutcomes}$$

Example 1
What is the probability you get a 6 when you roll a die?

A die has 6 sides, 1 side contains the number 6. We have 1 wanted outcome out of the 6 possible outcomes, therefore, the probability of getting a 6 is *1/6*.

a. Independent Events

Example 2
What is the probability of getting three 6s if we have 3 dice?

- The probability of getting a 6 on one die is *1/6*.
- The probability of getting three 6s is:
  $P(6,6,6) = 1/6 * 1/6 * 1/6 = 1/216$

This was an example of probability of Independent Events:

Two events are independent when the outcome of the first event does not influence the outcome of the second event — getting a 6 when rolling the first die does not affect the outcome of rolling the second die.

When we determine the probability of two independent events we multiply the probability of the first event by the probability of the second event:

```
P(X and Y) = P(X) * P(Y)
```

b. Dependent Events

Example 3
What is the probability of choosing two red cards in a deck of cards?

This is an example of Dependent Events:

Two events are dependent when the outcome of the first event affects the outcome of the second event. To determine the probability of two dependent events, we use the following formula:

```
P(X and Y) = P(X) * P(Y after X has occurred)
```

Getting back to our problem:

Since a deck of cards has 26 black cards and 26 red cards, the probability of randomly choosing a red card is:

$$P(red) = 26/52 = 1/2$$

However, now that we have already taken out one red card from the deck, the probability of choosing a red card ***again from the same deck*** becomes:

$$P(2_{nd}red) = 26/52 * 25/51 = 25/102$$

c. Mutually Exclusive Events

Two events are mutually exclusive when it is impossible for them to happen together. Turning left and turning right are mutually exclusive events — you can't do both at the same time. Similarly, getting heads and tails while tossing up a coin are also mutually exclusive. *Well, except in the world of quantum physicists!*

```
P(A and B) = 0
```

The probability that one of the events occurs is the sum of their individual probabilities.

```
P(X or Y) = P(X) + P(Y)
```

Example 4
a. What is the probability of getting a *King* and a *Queen* from a deck of cards?

A card cannot be a *King* AND a *Queen* at the same time! So the probability of a *King* and a *Queen* is 0 (impossible).

We can have a *King* OR a *Queen*. In a Deck of 52 Cards:

- the probability of a *King* is 1/13, so *P(King)=1/13*
- the probability of a *Queen* is also 1/13, so *P(Queen)=1/13*
- When we combine those two Events: P(*King* or *Queen*) = *(1/13)* + *(1/13)* = *2/13*

d. Inclusive Events

Inclusive events are events that can happen at the same time. To get the probability of an inclusive event, we first add the probabilities of the individual events and then subtract the probability of the two events occurring together:

```
P(X or Y) = P(X) + P(Y) - P(X and Y)
```

```
P(X or Y) = P(X) + P(Y) - P(X and Y)
```

Example 5
If you choose a card from a deck, what is the probability of getting a *Queen* or a *Heart*?

It is possible to get a *Queen* and a *Heart* at the same time, the *Queen of Hearts* which is the intersection *P(X and Y)*. So:

$$P(QueenOrHeart) = P(Queen) + P(Heart) - P(QueenOfHearts) = \\ = 4/52 + 13/52 - 1/52 = 16/52$$

Conditional Probability

Conditional probability is a measure of the probability of an event given that another event has occurred. In other words, it is the probability of one event occurring with some relationship to one or more other events.

Say event *X* is that it is raining outside, and there a 0.3 (30%) chance of rain today. Event *Y* might be that you will need to go outside with a probability of 0.5 (50%).

A conditional probability would look at these two events, *X* AND *Y*, in relationship with one another. In the previous example this would be the probability that it is both raining and you need to go outside.

Conditional probability is given by:

```
P(Y|X) = P(X and Y) / P(X)
```

Example 6
What is the probability of drawing 2 *Kings* from a deck of cards?

- For the first card the chance of drawing a *King* is 4 out of 52 since there are 4 *Kings* in a deck of 52 cards: *P(X) = 4/52*
- After removing a *King* from the deck, only 3 of the 51 cards left are *Kings*, meaning the probability of the 2nd card drawn being a King is less likely:*P(Y|X) = 3/51*
- So, the chance of getting 2 *Kings* is about 0.5%:

$$P(X and Y) = P(X) * P(Y|X) = (4/52) * (3/51) = 12/2652 = 1/221$$

A very important, and extensively used, derivation from conditional probability is the famous ***Bayes Theorem***. And that's what we are going to dive-in in the next lesson. The probability journey continues; see you in the next lesson!