Introduction to Probability

Probability - Why is it important?

The data generation process is highly uncertain, but understanding probability can help to deal with this uncertainty.

Data is always imperfect, and to draw inferences from this imperfect data, we treat the imperfections as the result of a random process.

Probability also helps us to understand how good our inferences are. Estimates will always have a probability attached or a margin of error, which can be determined by understanding probability.

Probability

Most processes are **stochastic** rather than **deterministic**. That is, the same inputs do not always produce the same outputs.

Why? The process could just be inherently random, or there are additional factors that we have not measured or have not considered that contribute to the different outcomes.

Probability

Deterministic Process: Finding the area of a circle based on its radius. As soon as we know the radius (r), we know the area (A) since there is an exact formula:

$$A = \pi r^2$$

Stochastic Process: Finding a man's shoe size based on his height. This is stochastic since it is not unusual for men who are exactly the same height to wear different sizes of shoe.

Probability - Informal Example

We can reason about random processes by using probability.

For example, let’s think about the possible outcome of a die roll using probability.

Probability - Informal Example

What is the chance of rolling a 5?

Answer: There are 6 equally-likely faces, and only one of them is 5, so the chances of rolling a 5 are 1/6.

Probability - Informal Example

What is the chance of rolling an even number?

Answer: There are 6 equally-likely faces, and three of them are even (2, 4, and 6), so the chances are 3/6.

Probability - Informal Example

What is the chance of rolling the die twice and it landing on a 4 both times?

Answer: We know that 1/6 of the time, it will land on 4 on the first roll and 1/6 of the time, it will and on 4 on the second roll.
Combining these, we can see that it will land on 4 both times $(1/6)*(1/6) = 1/36$ of the time.

Probability

Now, let's get a little more formal in talking about probability.

If you want to get very formal about it, there is a whole branch of mathematics dedicated to studying probability, starting from the framework of Kolmogorov's Axioms.

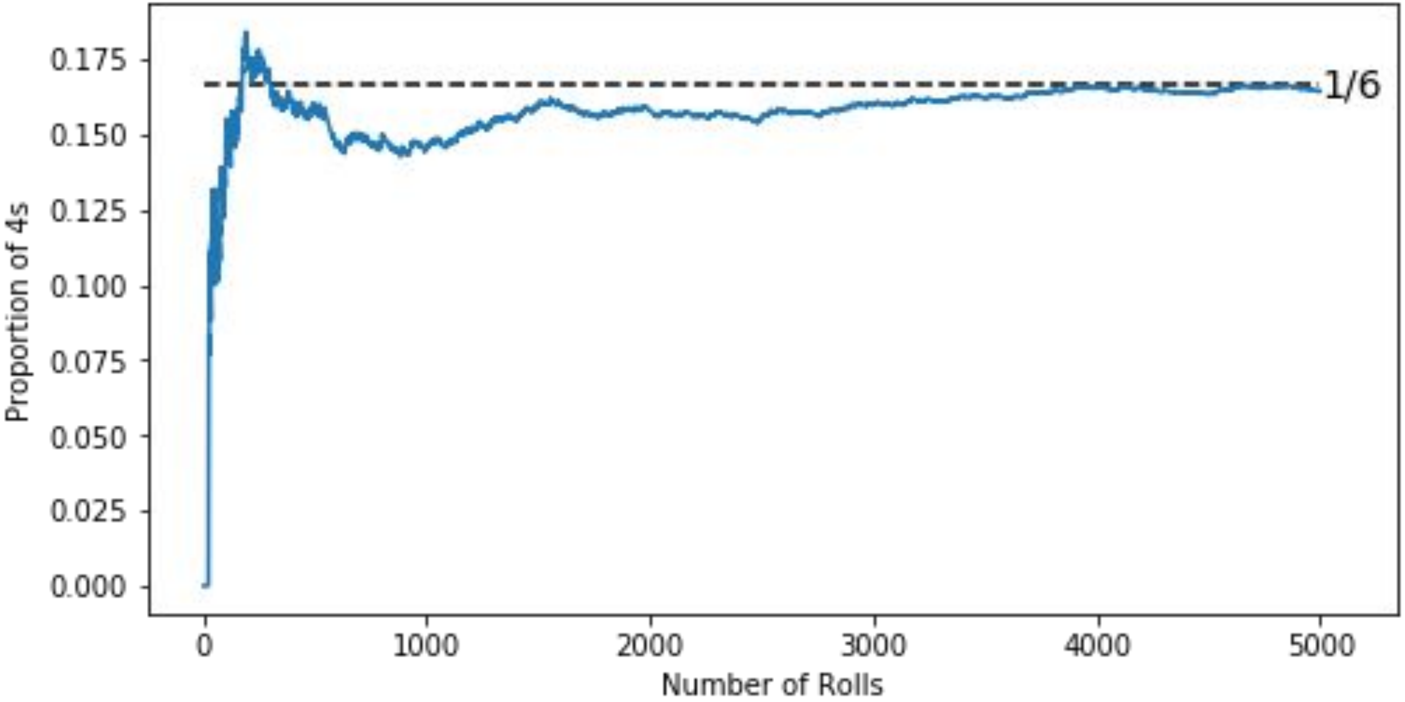
Probability as **Relative Frequency**

What do we mean when we say that the chance of rolling a 4 is 1/6?

Since there are 6 equally likely faces, if one were to roll the die a large number of times, then it should land on each face approximately 1/6 of the time.

If we continued to roll the die over and over and tracked how often it landed on 4, the proportion of times that the die lands on 4 should approach 1/6.

Probability as **Relative Frequency**



Probability as **Relative Frequency**

The probability of event A is the proportion of times that A occurs in a very long run of separate tries.

Law of Large Numbers: According to the law, the average of the results obtained from a large number of trials should be close to the expected value, and will tend to become closer to the expected value as more trials are performed.

This notion of probability is the **frequentist approach**.

Probability Terminology

In probability theory, we talk about the probability of an event.

Notation: For event A, the probability of A is written as $P(A)$

Probabilities range from 0 to 1, with 0 representing *impossibility* and 1 representing *certainty*.

An event with probability 0.7 will occur in about 70% of tries.

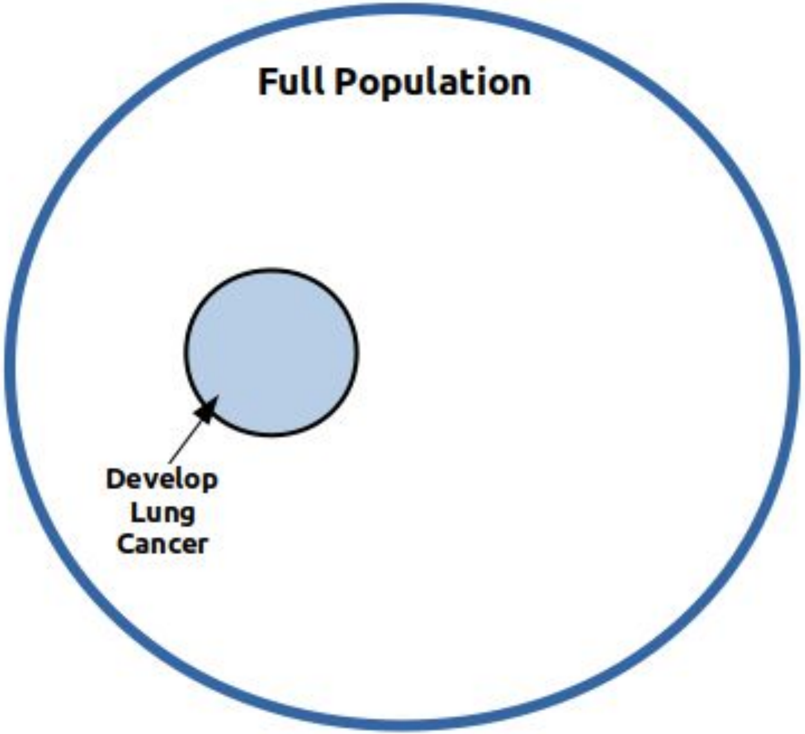
Conditional Probability

How does the likelihood of a particular outcome change, if we are given more information?

Eg. What is the probability that a person develops lung cancer?

What if we know that person smokes? How does the probability change?

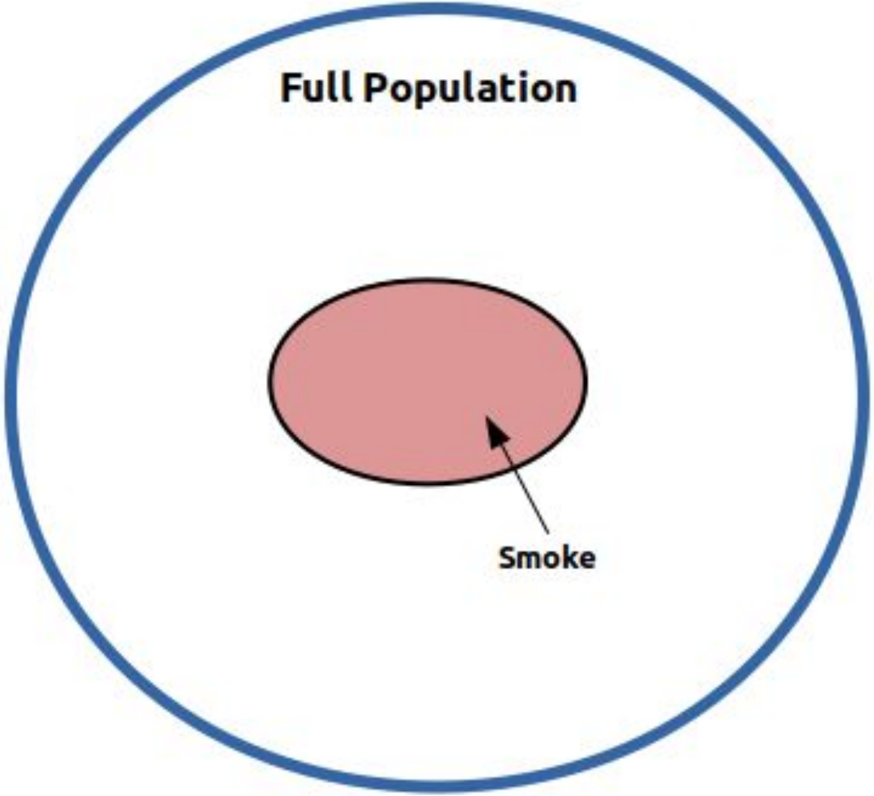
Conditional Probability



If we knew no other information about a person, our best guess at the likelihood of them developing lung cancer would be the ratio of people from the full population which develop lung cancer.

This starting probability is called the **prior probability.**

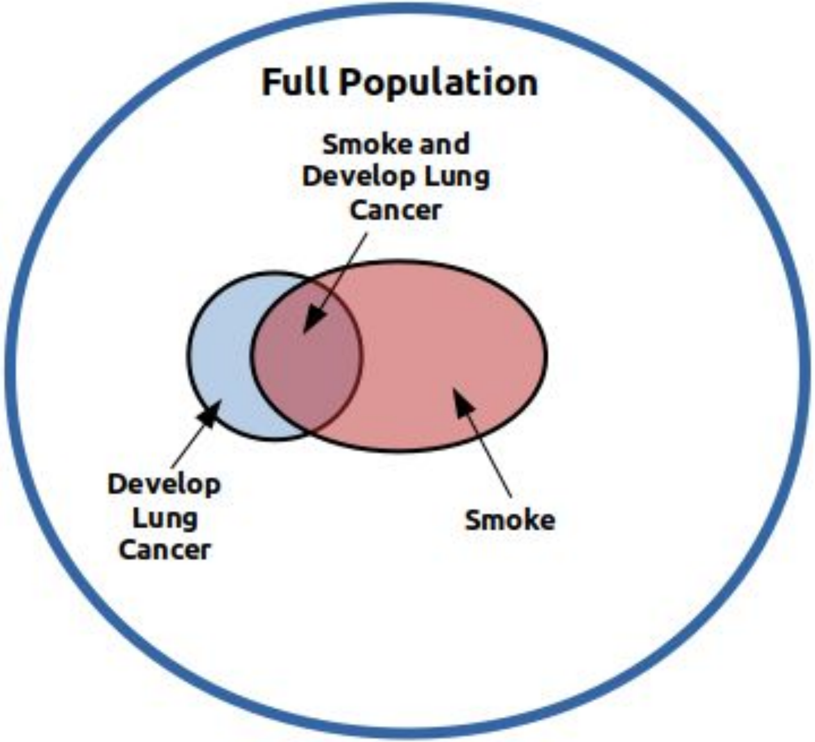
Conditional Probability



But if we had additional information, that they smoke, we could potentially use this information to refine our first guess.

To do this, we need to know the overlap between smokers and those who develop lung cancer.

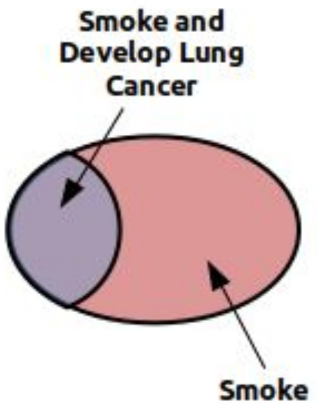
Conditional Probability



Let’s say that these two subpopulations overlap like this.

To update the probability that the smoker gets lung cancer, we no longer need to consider the full population, but instead just the subset of smokers.

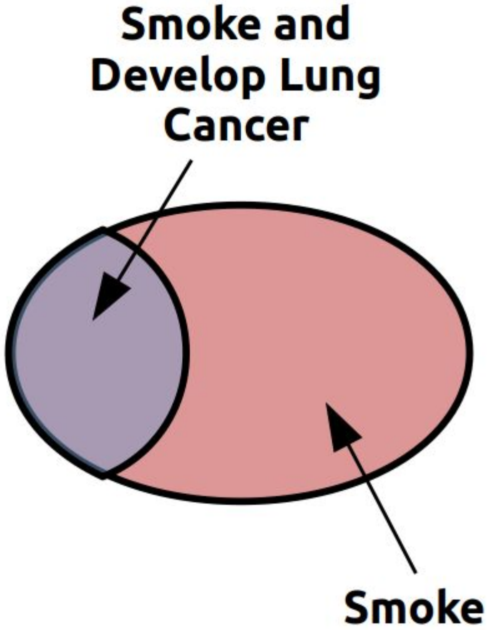
Conditional Probability



Let’s say that these two subpopulations overlap like this.

To update the probability that the smoker gets lung cancer, we no longer need to consider the full population, but instead just the subset of smokers.

Conditional Probability



That is, we can change our population on interest to just those people who smoke.

Now, our best guess for the likelihood of developing lung cancer (without any additional information) would be the ratio of people who smoke *and* develop lung cancer out of the population of smokers.

This is known as the **posterior probability.**

Conditional Probability

For two events *A* and *B*, the **conditional probability of *A*, given that *B* has occurred** is *P(A / B)*

Probability of developing lung cancer:

$$P(\text{lung cancer})$$

Probability of developing lung cancer, given that a person smokes:

$$P(\text{lung cancer} \mid \text{smokes})$$

Conditional Probability

To find P(A|B), we need to realize that we are looking for the proportion of outcomes where both *A* and *B* occur, but only need to consider those where *B* has occurred.

We can calculate the conditional probability as

$$P(A|B) = P(A \text{ and } B) / \mathrm{P}(B)$$

Probability as **Degree of Belief**

Thinking in terms of conditional probabilities, we can understand probability in a different way.

Probability quantifies how *certain* we are about a given hypothesis (event).

By incorporating more information, we can update our probabilities.

This interpretation of probability is known as the **Bayesian approach**.

Bayesian Statistics

The Bayesian approach is named after Thomas Bayes, an 18th century statistician, minister and philosopher.

Bayes formulated a special case of what is now known as *Bayes' Theorem*.

Prior

Bayes' Theorem

$$P(A \mid B) = \frac{P(B \mid A)}{P(B)} \cdot P(A)$$

Posterior

Bayes' Theorem can be viewed as a recipe for updating our belief about event A by incorporating information about event B.

Conditional Probability

Independent and Dependent Events

Events *A* and *B* are **independent** if the occurrence of *B* in no way informs us about the probability of *A*. The two events do not “interfere” with each other.

That is, P(A|B) = P(A)

Equivalently, P(A and B) = P(A)*P(B)

Otherwise, we way that *A* and *B* are **dependent**.

Conditional Probability

From the Bayesian point of view, if events A and B are independent, knowing that event A has occurred does not allow us to update our belief about the likelihood of B occurring.

Conditional Probability

Example: Flipping a coin twice

P(Second Coin Landing on Heads) = 0.5

P(Second Coin Landing on Heads | First Coin Landing on Heads) = 0.5

That is, knowing the outcome from the first flip gives us no additional information about the next one.

Conditional Probability

Why do we care?

When drawing a sample, we usually assume that the individuals are drawn *independently*.

When we're looking at a dataset, we usually assume that our observations are independent of each other.

Conditional Probability

We will usually make the further assumption that all of our observations arose from the same data generation process.

The combination of this assumption plus the assumption of independence is usually shortened to **iid**, standing for *independent and identically distributed*.