Probability theory is the mathematical field concerned with ***quantifying the likelihood*** of uncertain events.

- The *probability of an event* is a number between 0 and 1.
- The closer the number is to 1, the more certain we are that the event will occur.
- The closer the number is to 0, the more certain we are that the event will *not* occur.
- Events whose probabilities are close to 1/2 are those about which we are the most *uncertain*.
- The probability of an event depends on what we know (or assume). As the information changes, so can the probability.

The *frequentist* definition

The probability of an event is the ***relative frequency*** *with which the event will is observed if the same set of circumstances are repeated a large number of times. I.e., if the event E is observed k times in n repetitions, then the probability of E, denoted by $P(E)$, is k/n.*

The *classical* definition

If a process has n ***equally likely*** *outcomes and the event E comprises k of these outcomes, then $P(E) = k/n$.*

The frequentist and classical definitions are connected by...

The *law of large numbers*

If $P(E)$ is the (classical) probability of the event E, then as the number n of (independent) repetitions of the process grows large, the relative frequency with which E is observed gets closer and closer to $P(E)$ (almost certainly).

Example.

Suppose that a fair coin is tossed three times. The (eight) possible outcomes are

$$HHH, HHT, HTH, THH, TTH, THT, HTT, TTT$$

Unless we have information to the contrary, the only reasonable assumption is that these eight outcomes are equally likely.[a]

If E is the event *"exactly two H are observed in three tosses"*, then E is comprised of the outcomes HHT, HTH and THH, so

$$P(E) = \frac{\#\text{ of outcomes in } E}{\text{total }\#\text{ of outcomes}} = \frac{3}{8}.$$

The *law of large numbers* says in this case that if the process of tossing a coin three times is repeated a large number of times, then ***it is almost certain that*** the relative frequency of the event E (two H and one T) in this large number of repetitions ***will be very close to*** 3/8.

[a] An instance of Laplace's *principle of insufficient reason.*

Conditional probability

(*) The probability of an event depends on what we know (or assume). As the information changes, so can the probability.

(*) Suppose that E and F are two related events. The ***conditional probability*** of E given F, written $P(E|F)$, is the probability of observing event E, assuming that event F has been observed.

(*) If we know that the event F has occurred, then the set of possible outcomes has changed to include *only those* which occur in F.

(*) If $P(E|F) = P(E)$, i.e., if observing the event F does not change the probability of the event E, then we say that E and F are *statistically independent* (or just independent, for short).

Example. Returning to the three coin tosses, suppose that F is the event *"The first coin flip results in H"*, and E is as before ($2H$ and $1T$). If we know that F has occurred, then the set of possible outcomes is reduced to HHH, HHT, HTH and HTT. Two of these four outcomes result in E, so

$$P(E|F) = \frac{\#\text{ of outcomes in } E \text{ and in } F}{\#\text{ of outcomes in } F} = \frac{2}{4} = \frac{1}{2}.$$

Similarly, if we know that E has occurred, then the conditional probability of F given E is

$$P(F|E) = \frac{\#\text{ of outcomes in } E \text{ and in } F}{\%\text{ of outcomes in } E} = \frac{2}{3},$$

which is different (bigger) than the *unconditional* (or *prior*) probability of F,

$$P(F) = \frac{\#\text{ of outcomes in } F}{\text{total }\#\text{ of outcomes}} = \frac{4}{8} = \frac{1}{2}.$$

Looking at either case, we see that E and F are *dependent* (not *independent*).

Rules for calculating probabilities.

1. If the event E is certain ***not*** to occur, then $P(E) = 0$; if the event E is certain ***to*** occur, then $P(E) = 1$ and in any case $0 \leq P(E) \leq 1$.

2. The conditional probability of E given F is calculated as follows:

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

where "E and F" is the event where both E and F occur. Likewise,

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

In many cases, the conditional probability is known, and we use the formula(s) above to find the probability of "E and F":

$$P(E \text{ and } F) = P(E) \cdot P(F|E) = P(F) \cdot P(E|F).$$

This is called the *multiplication rule*. In the special case where E and F are *independent*, the multiplication rule gives

$$P(E \text{ and } F) = P(E)P(F).$$

3. The probability of "E or F" is given by

$$P(E \text{ or } F) = P(E) + P(F) - P(E \text{ and } F).$$

This is called the *addition rule*.

(*) If $P(E \text{ and } F) = 0$, the events E and F are called *mutually exclusive*. If E and F are mutually exclusive, then the addition rule simplifies to

$$P(E \text{ or } F) = P(E) + P(F).$$

(*) The event *"E does **not** occur"* (called the *complement* of E) is denoted by $\overline{E}$. Since it is certain that E either occurs or does not, and since E and $\overline{E}$ are mutually exclusive, the addition rules says

$$P(E) + P(\overline{E}) = P(E \text{ or } \overline{E}) = 1$$

which we often use to find $P(E) \implies P(E) = 1 - P(\overline{E})$.

(*) More generally if A and E are any events, then first of all,

$$A = (A \text{ and } E) \text{ or } (A \text{ and } \overline{E})$$

and second, the events $(A \text{ and } E)$ and $(A \text{ and } \overline{E})$ are mutually exclusive, so

$$P(A) = P(A \text{ and } E) + P(A \text{ and } \overline{E}).$$

Combining this with the multiplication rule we have the useful formula

$$P(A) = P(E)P(A|E) + P(\overline{E})P(A|\overline{E}).$$

4. ***Bayes' rule.*** Suppose that we know $P(B|A)$, how can we find $P(A|B)$? This question arises in situations where the natural sequence of events is *first* A*, then* B. In these situations we often know (or think we know) the *prior* (unconditional) probability $P(A)$, as well as the conditional probabilities $P(B|A)$ and $P(B|\overline{A})$. If all of this is known, then

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

Example. Suppose that a test for a certain dread disease (dd) has a probability of 0.02 of returning a *false positive* and a probability of 0.001 of returning a *false negative*. Furthermore, suppose that it is known that 1% of the population is infected with dd.

Question: If a random individual tests positive for dd, what is the probability that he or she is actually infected?

Answer: First, some notation. I'll use D to denote the event that the individual is infected with dd and T for the event that the individual tests positive for dd.

What we know is (i) $P(D) = 0.01$ (probability that a randomly individual is infected), (ii) $P(T|\overline{D}) = 0.02$ (probability of a false positive) and (iii) $P(\overline{T}|D) = 0.001$ (probability of a false negative). What we want to know is $P(D|T)$, and we can use Bayes' rule to find it.

First, $P(\overline{D}) = 1 - P(D) = 0.99$. Likewise

$$P(T|D) = 1 - P(\overline{T}|D) = 1 - 0.001 = 0.999$$

(conditional probabilities behave exactly like unconditional probabilities).

Now we can simply plug everything into Bayes' formula

$$\begin{aligned} P(D|T) &= \frac{P(T \text{ and } D)}{P(T)} \\ &= \frac{P(D)P(T|D)}{P(D)P(T|D) + P(\overline{D})P(T|\overline{D})} \\ &= \frac{(0.01)(0.999)}{(0.01)(0.999) + (0.99)(0.02)} = \frac{0.00999}{0.02979} \approx 0.335. \end{aligned}$$

Conclusion: Even though there is a very low probability of a false positive, a positive test result indicates an infected person in only about 1 out of 3 cases. This may appear contradictory at first, but you have to remember that a false positive is the conditional event *"positive result given no infection"* $(T|\overline{D})$, while the event whose probability we just calculated is *"infection given positive test"* $(D|T)$.

In this example the probability $P(D|T)$ is relatively low because of the *prior* probability $P(D) = 0.01$. Most people are not infected, so even though very few of them test positive, almost two thirds of all positive results (in this hypothetical example) are false positives.