Introduction to Probability

In casual conversation, people tend to use the concepts of probability quite loosely. We are going to give precise definitions and formulas to give probability theory a firm footing in discrete mathematics.

Sample Spaces

Probability theory is based on the idea of observing the outcome of an experiment that has different possible outcomes. We call this process of experimenting and observing ***sampling***. The type of experiment we are usually talking about here is one with a fixed set of possible outcomes, such as flipping a coin, tossing a six-sided die, or picking one ball out of a box of different coloured balls.

A ***sample space*** consists of the set of possible outcomes of an experiment, and a function P( ) that assigns a value to each outcome. P( ) must have the following properties:

$0 \leq P(x) \leq 1$ for each outcome x

$$\sum_{x \in S} P(x) = 1$$

We call P( ) a probability function.

We will sometimes write (S,P) to identify the sample space where S is the set of outcomes and P is the probability function.

Any function that satisfies the requirements is a valid probability function, but we usually want our probability functions to correspond to the "likelihood" of the different outcomes occurring. But that is dangerously close to a circular definition, since "likelihood" is often used as a synonym for "probability".

We can get a concrete sense of what we want our probability functions to do by considering an experiment which we sample many, many times. If we count the number of times a particular outcome occurs and divide that by the number of samples, we expect that this ratio will change less and less as the number of samples increases. The limit of this ratio as the number of samples goes to ∞ is what we want as the probability of that particular outcome. However we want to want to do this without working out limits because that's calculus!

For example, consider a box containing a red ball, a yellow ball and a blue ball. If the balls are all identical in size and weight, we **expect** that if we take out one ball, record its colour and return it to the box, over and over again, the number of times we withdraw the red ball divided by the total number of samples, will get closer and closer to 1/3 ... as will the ratios for the yellow ball and the blue ball.

Now suppose the box contains 2 red balls and 1 yellow ball. The possible outcomes from our sampling experiment are {red, yellow}. Again assuming that the balls are identical except for their colours, we **expect** that the ratio for red (the ratio of occurrences / samples) will approach 2/3, while the ratio for yellow approaches 1/3.

So in general, we want P(x) to be the ratio "occurrences of x"/"number of samples" as the number of samples goes towards ∞

It's important to discuss why I put the word "**expect**" in bold in the two paragraphs above. This is the way we believe the world works: that if we toss a balanced coin many many many times, the number of "Heads" and "Tails" will be more or less equal. We accept the possibility that the coin will come up "Heads" every time (see the film "***Rosencrantz & Guildenstern Are Dead***") but we don't expect it to happen that way. Of course this is based on countless observations as well. The theory of probability that we are developing is designed to make this slightly vague "expectation of the way the world works" into a solid mathematical tool for describing and predicting aspects of the real world.

Events

If (S,P) is a sample space, we use the word **event** to describe any subset of S. For example, if S = {1,2,3,4,5} then A = {1,3,5} is an event. {} is also an event, and so is S.

Suppose we sample (S,P) by conducting the experiment once. If the outcome of the sample is an element of A, we say that **A has occurred**.

Now we can define the **probability of an event**: the probability of event A is the sum of the probabilities of the elements of A. In notation

$$P(A) = \sum_{a \in A} P(a)$$

For example, given S as above, suppose P(1) = P(2) = P(3) = P(4) = P(5) = 0.2

Then P({1,3,5}) = 0.6

But suppose P(1) = 0.4, P(2) = 0.1, P(3) = 0.3, P(4) = 0.1, P(5) = 0.1

Then P({1,3,5}) = 0.8

In casual discussions of probability, people often forget that the probability of an event depends on the probability function – they assume (without cause) that all outcomes are equally probable. We often hear people say things like "It could go either way, so it's 50/50" ... meaning that because there are two possible outcomes, both have equal probability. But there is no reason to assume that both outcomes have equal probability. Suppose I have bag containing a very large cow and a very small mouse, and I reach in and pull out one of the animals. There are only two possible outcomes, but we expect that I will have a cow (man) more often than a mouse. Another very simple example is when the experiment consists of rolling two fair 6-sided dice and adding the numbers that come up. The possible outcomes are {2,3,4,...,12} but the probabilities are not all equal. So the probability of the event {6,7,8} is very different from the probability of the event {2,3,4} even though both contain the same number of outcomes.

To be honest, the (unjustified) assumption that all outcomes of an experiment are equally probable sometimes shows up in scientific discussions as well.

Combinations of Events

Let (S,P) be a sample space, and let A and B be events in that sample space.

What can we say about $A \cup B$? More precisely, can we compute $P(A \cup B)$ from P(A) and P(B)?

Unfortunately P(A) and P(B) do not give enough information to compute $P(A \cup B)$.

Consider this example. Let the experiment be tossing a single 6-sided die. We will assume that all outcomes are equally probable (i.e. $P(i) = \frac{1}{6}\ \forall\ i \in \{1,2,3,4,5,6\}$)

Let $A = \{1,2,3\}$ $B = \{2,3,4\}$ $C = \{4,5,6\}$

$P(A) = P(B) = P(C) = \frac{1}{2}$, but $P(A \cup B) = P(\{1,2,3,4\}) = \frac{2}{3}$ and $P(A \cup C) = 1$

In the first case two events with individual probabilities $= \frac{1}{2}$ have a combined probability $= \frac{2}{3}$, and in the second case two events with individual probabilities $= \frac{1}{2}$ have a combined probability = 1

The difference of course is that A and B overlap (i.e. they have non-empty intersection) while A and C do not overlap ... and we can't tell that just by looking at their probabilities. Fortunately the solution is obvious as soon as we recognize the problem. We just apply our old friend the Principle of Inclusion/Exclusion, and arrive at this formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

Note that this means the only time $P(A \cup B) = P(A) + P(B)$ is when $P(A \cap B) = 0$

Also observe that using this equation, if we know any three of the terms we can deduce the fourth. For example if we know $P(A \cup B) = 0.7$, $P(A) = 0.3$, $P(A \cap B) = 0.2$, then we know $P(B) = 0.6$

Here are some more useful facts about the probabilities of events

$P(\emptyset) = 0$

$P(S) = 1$

$P(A \cup B) \leq P(A) + P(B)$

$P(\overline{A}) = 1 - P(A)$ where $\overline{A}$ represents the complement of A