Probability: Events

Introduction to Probability: Events

Randomness plays an important part in statistics. Sometimes randomness is a feature of what we observe in our data, for example noise during measurement process. Often researchers deliberately introduce randomness through random sampling or the random assignment of treatments to subjects in order to be able to make better conclusions from their studies. So understanding some basics of probability is important.

We have experience with randomness everyday:

Examples

- Flip a coin and get Heads or Tails
- The song that comes up next when playing music on shuffle
- The number of each colour that are in a handful grabbed from a box of coloured candies
- How many Conservative party supporters are in a randomly chosen sample of people who are polled
- The sex of the next baby born at a hospital
- The birth weight of the next baby born at a hospital

In every day conversation, people often use random to mean haphazard or a variety of other things, but to a statistician it has a more definite meaning. To a statistician, an experiment is random if individual outcomes of it are uncertain (i.e. we don't know what the outcome will be), but if we repeat the experiment a large number of times, there is a structure to how often outcomes occur.

Random: *Individual outcomes are uncertain, but there is a structure to how often outcomes occur in a very large number of repetitions.*

So although we don't know what might happen in a particular experiment, the fact that we know what to expect in the long term (how often we will get heads or how often we will get a large baby over 4500 grams) allows us to create a mathematical model for the randomness which is a probability distribution for the random experiment. The probability distribution allows us to quantify the underlying structure that exists in the long run.

Probability distributions: *are mathematical models to describe what happens in the long run with outcomes of random experiments.*

Example

Consider flipping a coin once. We don't know if we will get H or T on any one flip, but if we flipped our coin many times we would expect to get close to 50% Heads and 50% Tails. Hence the probability distribution for the random experiment of a coin flip is that the probability of Heads is 0.5 and the probability of Tails is 0.5.

This kind of interpretation of probability as the proportion of times something will happen in the long run is a common and often very useful way to think of probabilities.
It is also possible to base probabilities on theoretical arguments. Such as designing a die to be equally likely to come up on all sides, so there exist physical arguments for the fact that each side is equally likely to come up when we roll the die.
Probabilities are also often subjective (based on our opinion) which hopefully has some good reasons to support. For example, you might have a reasoned opinion for the probability that your favourite sports team will make the playoffs next year.

For what we will do going forward, the *long run relative frequency* interpretation will work, so you can think about probabilities that way.

We're usually interested in the probability of some event, which specifies some subset of the possible outcomes of our experiment.

Event: *A subset of the possible outcomes of a random experiment.*

Examples of Events

- Getting six heads when we flip a coin 10 times
- Getting a song from the 90's as the next song in a random shuffle
- Getting a majority of Conservative party supporters in a poll
- Getting a baby that weighs more than 4500 g

Events are usually given capital letters, early in the alphabet. So we can define:

- A = Getting six heads
- B = Getting a song from the 90's
- C = Getting a majority of Conservative party supporters
- D = Getting a baby that weighs more than 4500 g

If, in general, E is some event from a random experiment, we write

$P(\mathrm{E})$ = the probability of the event E

To be a valid probability, $P(\mathrm{E})$ must be a number between 0 and 1.
If $P(\mathrm{E}) = 1$, E is sure to occur
If $P(\mathrm{E}) = 0$, E is sure not to occur.

The complement of an event, written E^c is the event that E does not happen.
Also

$$P(E^c) = 1 - P(E)$$

EXAMPLE
If the probability that the next song on our random shuffle is from the 90's is 0.3, then the probability that it is from another decade is 0.7.

$P(\text{next song is not from the 90's}) = 1 - P(\text{next song is from the 90's}) = 1 - 0.3 = 0.7$