NCEA Probability

Alex Elzenaar

March 10, 2019

0 Preliminaries: sets

A set is a collection of objects. We will write a set by either listing its objects inside curly braces; so the set of the first three whole numbers is $\{1, 2, 3\}$, or by giving a rule that can be used to decide whether a given object is in the set or not, like $\{p : \text{ where } p \text{ is a prime}\}$. Two sets are equal if (and only if) they contain the same elements (objects, members). The order of elements in a set does not matter, and elements can only occur once in a set (so $\{1, 2, 3\} = \{1, 2, 3, 2\} = \{3, 2, 1\}$). If A is a set, the number of distinct elements in A is denoted by $|A|$.

The set with no elements is called the *empty set*, denoted $\emptyset$. Caution: $\emptyset \neq \{\emptyset\}$!

If A and B are two sets, then the union $A \cup B$ is the set consisting of every object that is in either A or B (or both); so $\{1, 2\} \cup \{\gamma, \omega\} = \{1, 2, \gamma, \omega\}$. The intersection $A \cap B$ is the set of all elements that are in both A and B; so $\{1, 2, 3\} \cap \{3, 4, 5\} = \{3\}$, and $\{\text{odd numbers}\} \cap \{\text{even numbers}\} = \emptyset$.

If every element of A is also in B, then A is called a subset of B and we write $A \subseteq B$.

1 Random processes

Suppose we have some process that produces a set of outcomes, such that even when the same inputs are given to the process different outputs might occur. We will call such a process a *random process*. We will denote random processes by names in sans-serif fonts, like Rand.

Random processes occur naturally in a variety of situations.

1.1 Example.

1. When we toss an ideal coin, there are two different outcomes that may occur, and given the same initial conditions both outcomes are possible and both will occur as we toss the coin multiple times. Thus we can model the coin toss as a random process Coin.

2. In cryptography, one of the major components of any encryption/decryption system is a process that produces a large prime number that should be unknown to any attacker (or to anyone who should not be able to decrypt the data). Such a process should be a random process.

One of the fundamental things we would like to be able to do is predict 'how likely' a random process is to produce a given output. To do this, we will need to model outputs of random processes.

1.2 Definition. Suppose Rand is a random process. The set $\Omega(\mathsf{Rand})$ of possible outcomes is called the *sample space* of the process. The size$|\Omega\mathsf{Rand}|$ of the sample space is denoted $N(\mathsf{Rand})$.

For the time being, we are only interested in random processes Rand such that $N(\mathsf{Rand}) < \infty$ — i.e. such that the sample space is finite.

1.3 Example.

1. For the random process Coin of tossing an ideal coin, the sample space is $\Omega(\mathsf{Coin}) = \{H, T\}$.
2. For the random process of tossing an ideal coin three times, the sample space is $\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$.
3. For the random process of tossing a coin three times and counting the number of tails, the sample space is $\{0, 1, 2, 3\}$.
4. Define a random process as follows: toss a coin; if the result is H then toss a six-sided die; otherwise if the result is T flip the coin again. Then the sample space is $\{H1, H2, H3, H4, H5, H6, TH, TT\}$.

1.4 Exercise. Consider a bag with five coloured balls, three red and two black. Let Pick be the random process "take a ball from the bag without looking, note the colour, replace it, and pick out another ball and note the colour". What is the sample space of the random process if (a) the order matters, and if (b) the order doesn't.

Note that in the previous examples we are very interested in performing two random processes one after the other, and that sometimes a random process may depend on the result of a previous process.

In reality, we are less interested in the particular outcomes, and more interested in whether an outcome lies in a particular subset of the sample space: if a die is thrown, is the result even?

1.5 Definition. If Rand is a random process, then an *event* E is a subset of $\Omega(\mathsf{Rand})$ such that, under the conditions of the random process, it is possible to decide whether a given output is in E.

1.6 Example. Let us toss a coin three times; the sample space is

$$\{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\},$$

and the set

$$A = \{HHT, HTH, THH, HHH\}$$

is the event "at least two heads are seen".

Now consider the following experiments:

1.7 Example.

1. We toss a coin five times; we are interested in the event "the first toss is heads *or* the third toss is heads (or both)".
2. We roll two dice simultaneously, and we are interested in the event "both dice roll a six".
3. We roll six dice, and we are interested in the event "at least three rolls are even, but the third dice comes up odd".