UNIVERSITY OF LONDON

Probability and Statistics: To p, or not to p?

Module Leader: Dr James Abdey

2.2 Simple probability distributions

One can view probability as a quantifiable measure of one's degree of belief in a particular **event**, or **set**, of interest. Let us consider two simple experiments.

Example

i. The toss of a (fair) coin: $S = \{H, T\}$, where H and T denote 'heads' and 'tails', respectively, and are called the **elements** or **members** of the sample space.

ii. The score of a (fair) die: $S = \{1, 2, 3, 4, 5, 6\}$.

So the coin toss sample space has two elementary outcomes, H and T, while the score on a die has six elementary outcomes. These individual elementary outcomes are themselves events, but we may wish to consider slightly more exciting events of interest. For example, for the score on a die, we may be interested in the event of obtaining an even score, or a score greater than 4 etc. Hence we proceed to define an event of interest.

Typically, we can denote events by letters for notational efficiency. For example, $A =$ 'an even score' and $B =$ 'a score greater than 4'. Hence $A = \{2, 4, 6\}$ and $B = \{5, 6\}$.

The universal convention is that we define probability to lie on a scale from 0 to 1 inclusive.[1] Hence the **probability of any event A**, say, is denoted $\boldsymbol{P(A)}$ and is a real number somewhere in the unit interval, i.e. $P(A) \in [0, 1]$, where '$\in$' means 'is a member of'. Note the following.

- If A is an **impossible event**, then $P(A) = 0$.
- If A is a **certain event**, then $P(A) = 1$.
- For events A and B, if $P(A) > P(B)$, then A is more likely to occur than B.

[1] Multiplying by 100 yields a probability as a percentage.

Therefore, we have a probability scale from 0 to 1 on which we are able to *rank* events, as evident from the $P(A) > P(B)$ result above. However, we need to consider how best to *quantify* these probabilities theoretically (we have previously considered determining probabilities subjectively and by experimentation).

Let us begin with experiments where each elementary outcome is **equally likely**, hence our (fair) coin toss and (fair) die score fulfil this criterion (conveniently).

Determining event probabilites for equally likely elementary outcomes

Classical probability is a simple special case where values of probabilities can be found by just counting outcomes. This requires that:

- the sample space contains only a *finite* number of outcomes, N
- all of the outcomes are *equally probable* (*equally likely*).

Standard illustrations of classical probability are devices used in games of chance:

- tossing a fair coin (heads or tails) one or more times
- rolling one or more fair dice (each scored 1, 2, 3, 4, 5 or 6)
- drawing one or more playing cards at random from a deck of 52 cards.

We will use these often, not because they are particularly important but because they provide simple examples for illustrating various results in probability.

Suppose that the sample space S contains N equally likely outcomes, and that event A consists of $n \leq N$ of these outcomes. We then have that:

$$P(A) = \frac{n}{N} = \frac{\text{number of outcomes in } A}{\text{total number of outcomes in the sample space } S}.$$

That is, the probability of A is the **proportion** of outcomes which belong to A out of all possible outcomes.

In the classical case, the probability of any event can be determined by **counting** the number of outcomes which belong to the event, and the total number of possible outcomes.

Example

i. For the coin toss, if A is the event 'heads', then $N = 2$ (H and T) and $n = 1$ (H). So, for a *fair*[2] coin, $P(A) = 1/2 = 0.5$.[3]

[2] Remember we are assuming equally likely elementary outcomes here, so a fair coin is required. If we had a *biased* coin, then this approach would fail to accurately quantify probabilities.

[3] Probabilities can be reported as proper fractions or in decimal form.

ii. For the die score, if A is the event 'an even score', then $N = 6$ (1, 2, 3, 4, 5 and 6) and $n = 3$ (2, 4 and 6). So, for a *fair* die, $P(A) = 3/6 = 1/2 = 0.5$.

   Finally, if B is the event 'score greater than 4', then $N = 6$ (as before) and $n = 2$ (5 and 6). Hence $P(B) = 2/6 = 1/3$.

Example

Rolling two dice, what is the probability that the sum of the two scores is 5?

- Determine the sample space, which is the 36 ordered pairs:

$$\begin{aligned} S = \{&(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),\\ &(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),\\ &(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),\\ &(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),\\ &(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),\\ &(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)\}. \end{aligned}$$

- Determine the outcomes in the event $A = \{(1,4),(2,3),(3,2),(4,1)\}$ (highlighted).
- Determine the probability to be $P(A) = 4/36 = 1/9$.

Random variables

A **random variable** is a 'mapping' of the elementary outcomes in the sample space to real numbers. This allows us to attach probabilities to the experimental outcomes. Hence the concept of a random variable is that of a measurement which takes a particular value for each possible trial (experiment). Frequently, this will be a numerical value.

Example

Suppose we sample at random five people and measure their heights, hence 'height' is the random variable and the five (observed) values of this random variable are the realised measurements for the heights of these five people.

Suppose a fair die is thrown four times and we observe two 6s, a 3 and a 1. The random variable is the 'score on the die', and for these four trials it takes the values 6, 6, 3 and 1. (In this case, since we do not know the true order in which the values occurred, we could also say that the results were 1, 6, 3 and 6 or 1, 3, 6 and 6, or ....)

An example of an experiment with non-numerical outcomes would be a coin toss, for which recall $S = \{H, T\}$. We can use a random variable, X, to convert the sample space elements to real numbers such as:

$$X = \begin{cases} 1 & \text{if heads} \\ 0 & \text{if tails.} \end{cases}$$

The value of any of the above variables will typically vary from sample to sample, hence the name 'random variable'.

So each experimental random variable has a collection of possible outcomes, and a numerical value associated with each outcome. We have already encountered the term 'sample space' which here is the set of all possible numerical values of the random variable.

Example

Examples of random variables include the following:

Experiment	Random variable	Sample space
Die is thrown	Value on top face	$\{1, 2, 3, 4, 5, 6\}$
Coin is tossed five times	Number of heads	$\{0, 1, 2, 3, 4, 5\}$
Twenty people sampled	Number with blue eyes	$\{0, 1, 2, \ldots, 19, 20\}$
Machine operates for a day	Number of breakdowns	$\{0, 1, 2, \ldots\}$
One adult sampled	Height in cms.	$\{[150\text{cm}, 200\text{cm}]\}$ (roughly)

Probability distribution

A natural question to ask is 'what is the probability of any of these values?'. That is, we are interested in the **probability distribution** of the experimental random variable.

Be aware that random variables comes in two varieties – discrete and continuous.

- **Discrete:** Synonymous with 'count data', that is, random variables which take non-negative integer values, such as $0, 1, 2, \ldots$. For example, the number of heads in n coin tosses.

- **Continuous:** Synonymous with 'measured data' such as the real line, $\mathbb{R} = (-\infty, \infty)$, or some subset of $\mathbb{R}$, for example the unit interval $[0, 1]$. For example, the height of adults in centimetres.

The mathematical treatment of probability distributions depends on whether we are dealing with discrete or continuous random variables. We will tend to focus on discrete random variables for much of this course.

In most cases there will be a higher chance of the random variable taking some sample space values relative to others. Our objective is to express these chances using an associated probability distribution. In the discrete case, we can associate with each 'point' in the sample space a probability which represents the chance of the random variable being equal to that particular value. (The probability is typically non-zero, although sometimes we need to use a probability of zero to identify impossible events.)

To summarise, a probability distribution is the complete set of sample space values with their associated probabilities which must sum to 1 for discrete random variables. The probability distribution can be represented diagrammatically by plotting the probabilities against sample space values.

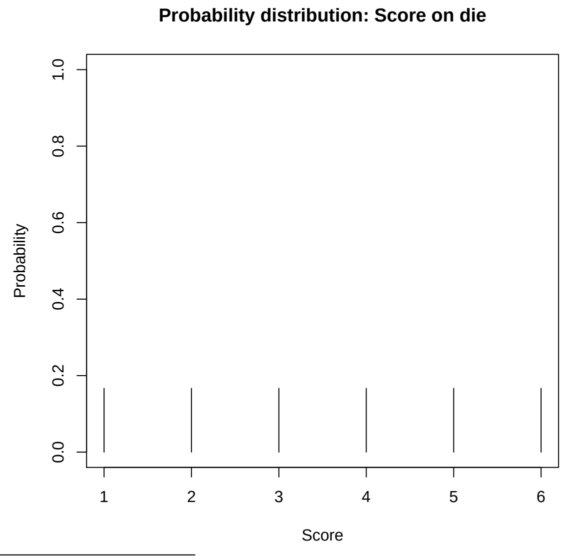
Finally, before we proceed, let us spend a moment to briefly discuss some important issues with regard to the notation associated with random variables. For notational efficiency reasons, we often use a *capital* letter to represent the random variable. The letter X is often adopted, but it is perfectly legitimate to use any other letter: Y, Z etc. In contrast, a *lower case* letter denotes a particular *value* of the random variable.

Example

Let X = 'the upper-faced score after rolling a fair die'. If the die results in a 3, then this is written as $x = 3$. The probability distribution of X is:

$X = x$	1	2	3	4	5	6
$P(X = x)$	1/6	1/6	1/6	1/6	1/6	1/6

This is an example of the **(discrete) uniform distribution**.[4] For discrete random variables, we talk about a **mass of probability** at each respective sample space value. In the discrete uniform case this mass is the same, i.e. 1/6, and this is plotted to show the probability distribution of X, as shown below.



[4]At school, 'uniforms' are worn, i.e. all pupils wear the *same* uniform, hence when the term 'uniform' is applied to a probability distribution, we have the *same* probability of occurrence for each sample space value.