16 Probability

16.1 Probability distributions

- Probability is, roughly, concerned with the *frequency* of *events.* It involves *taking samples* or *making trials* (probably synonymous) whose outcomes vary but follow some well-known *probability distributions.*

- A particular 'trial' should have several possible outcomes. For example: tossing a coin has two outcomes, 'heads' and 'tails.' Throwing two dice has 11 outcomes, the numbers 2 to 12.

  The set of possible outcomes is called a *sample space.*

- A probability distribution is a map $x_i \mapsto p_i$, $1 \leqslant i \leqslant k$ (k is the number of outcomes), where $0 \leqslant p_i \leqslant 1$, and $\{x_1, \ldots, x_k\}$ is the sample space.[1]

- Examples of distributions: uniform, binomial, Poisson, normal, Student's t-distribution, chi-squared.

- Imagine a large number of n independent trials.[2] Let n_j be the number of times the outcome x_j happens ($1 \leqslant j \leqslant k$, where k is the size of the sample space), so $\sum_j n_j = n$.

- The trials *follow the given distribution* if

$$\frac{n_j}{n} \approx p_j, \quad 1 \leqslant j \leqslant k$$

  and the larger n is the better the approximation will be.

- This is a bit woolly. It should be called the 'frequentist principle.'

(16.1) Corollary *(to the frequentist principle).*

$$0 \leqslant p_j \leqslant 1, \quad (j = 1, \ldots, k)$$
$$\sum_j p_j = 1$$

(16.2) Uniform distribution. *All distributions* are ultimately derived from the *uniform distribution*, in which all outcomes are equally probable:

$$p_i = \frac{1}{k}, \quad 1 \leqslant i \leqslant k.$$

(16.3) A fair coin. In tossing a coin, there are two outcomes H, T (heads, tails). In a long run of coin-tosses, unless something is wrong, there is no reason for one outcome to happen significantly more than the other. So we expect

$$\frac{n_1}{n} \approx \frac{n_2}{n} \approx \frac{1}{2}.$$

[1] Actually the sample space could be infinite (Poisson distribution) or continuous (Normal distribution).
[2] This means that the outcome of one trial has no effect on any other trial. A precise definition is possible.

(n_1 is the number of heads, n_2 tails, and $n = n_1 + n_2$). If this doesn't happen with large n then we say the coin is *biased,* otherwise it is *fair.*[3]

16.2 Groups of trials

If we consider sets of 3 coin-tosses, the outcomes can be

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

These groups of trials also come with a probability distribution.

(16.4) Proposition (multiplication principle). *Given m independent trials from a particular distribution, the probability of a particular* ***compound*** *outcome, i.e., a sequence of m particular outcomes,*

$$x_{i_1}\ x_{i_2}\ \dots\ x_{i_m} \quad \textit{is the product}$$
$$p_{i_1}\ p_{i_2}\ \dots\ p_{i_m}$$

For a fair coin, the probability of

$$HHH$$

is therefore $(1/2)^3 = 1/8$. The same for the other 7 outcomes. In other words, this is a sample space of size 8 with a uniform distribution.

(16.5) Throwing a single die (the plural is dice): the sample space has 6 outcomes, $1, \dots, 6$. If it is a fair die then the distribution is uniform.

Throwing two dice A and B has 36 outcomes,

$$(1,1), (1,2), \dots, (6,5), (6,6)$$

16.3 Sets of outcomes: events

(16.6) Definition *Let $S = \{x_1, \dots, x_k\}$ be a sample space (with associated probabilities p_j).*

An ***event*** *is 'simply' a subset of S.*

(16.7) Proposition (addition principle). *Given a sample space $S = \{x_1, \dots, x_k\}$ with probability distribution $\{p_1, \dots, p_k\}$, let $E = \{x_{i_1}, \dots, x_{i_m}\}$ be an event.*

Then the probability that a particular trial will have one of the outcomes which belong to E is the sum

$$p_{i_1} + \dots + p_{i_m}.$$

(This follows from the frequentist principle).

For example, if we consider the outcomes

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

and classify them by the number of heads, we get

[3] The probability of getting 10 heads in a row is about 0.1%. Yet Darren Brown did this on television. How?

- 3 heads: HHH, probability 1/8.
- 2 heads: HHT, HTH, THH, probability 3/8.
- 1 head: HTT, THT, TTH, probability 3/8.
- No heads: TTT, probability 1/8.

Of course this is leading towards the *binomial distribution.*

16.4 Balls in jar and Binomial Distribution

Suppose that one is given a number of red and green billiard balls. They are placed in a jar. There are b balls in total. A single trial involves taking a ball from the jar without looking, and replacing it in the jar. Assuming a uniform distribution, each ball can be extracted with probability $1/b$.

Suppose there are r red balls and g green, with $b = r + g$. Under the addition principle, the probability of taking out a red ball is $r/(r+g)$ and the probability of a green ball is $g/(r+g)$.

Suppose there are 6 red balls and 3 green in the jar. Let x_1 be the outcome 'red' and x_2 the outcome 'green.' So $p_1 = 2/3$ and $p_2 = 1/3$. Again, if we take groups of 3 trials, R for red and G for green, there are 8 outcomes

$$RRR, RRG, RGR, RGG, GRR, GRG, GGR, GGG.$$

But the probabilities of these outcomes is not uniform: respectively,

$$8/27,\ 4/27,\ 4/27,\ 2/27,\ 4/27,\ 2/27,\ 2/27,\ 1/27$$

If we classify these according to the number of red balls,

- 3 red: 8/27
- 2 red: 12/27
- 1 red: 6/27
- no red: 1/27

These are examples of the Binomial Distribution. A *Bernoulli trial* is one with 2 outcomes, A, B, say, where A has probability p and B has probability $1-p$. Given a group of n independent trials, the probability p_i of A occurring i times, $0 \leqslant i \leqslant n$, is

$$p_i = \binom{n}{i} p^i (1-p)^{n-i}.$$

This distribution is denoted $B(n,p)$.

The distributions considered here were $B(3, 1/2)$ and $B(3, 2/3)$.