

Probability summary

Definitions

Probability means "what proportion of the time do we expect a particular result from a random process". It is given a number between 0 (never happens) and 1 (certain to always happen).

If we add the probability of every possible result, the total probability must be 1. (There are no other alternatives).

Individual random events are unpredictable. I cannot predict tonight's lottery numbers. What I can predict, though, is the average of a large number of experiments. (e.g. for the lottery I would predict that each number is equally likely - and tables of thousands of past winning numbers show this is true).

The result of an experiment is called an "outcome" (e.g. a head or a tail).

The posh word for an experiment is a "trial".

Ways of finding a probability

1. If you have a number of equally likely outcomes, you can find the probability exactly.

$$\text{probability} = \frac{\text{number of "successful" outcomes}}{\text{number of possible outcomes}}$$

e.g. the probability of throwing a 1 or a 2 on a dice, you have 2 ways of being successful out of 6 possibilities, each face has a probability of $\frac{1}{6}$ so $P(1 \text{ or } 2) = \frac{2}{6} = \frac{1}{3}$

2. If you cannot use "equally likely outcomes" you need to do a series of experiments ("trials"). These give you a *relative frequency* that is an *estimate of the probability*.

The more trials you do, the better the estimate.

$$\text{relative frequency} = \text{estimated probability} = \frac{\text{number of "successful" outcomes}}{\text{number of trials}}$$

3. If you cannot do an experiment, you need to do the same calculation but using data that already exists, e.g. aircraft accident data for the last 10 years.

4. Often, politicians will just guess a probability. This can be wildly inaccurate (e.g. the space shuttle was thought to have 1 in 10,000 risk of disaster when it was probably about 1 in 100).

Relative frequency examples.

Throwing dice, using Excel as a random number generator (=ROUND(RAND()*6+0.5, 0) will generate an integer in the range 1 to 6). From one experiment, I got:

Number of throws	6	24	120	6000
Number of ones	0	3	17	972
Number of twos	1	3	27	1054
Number of threes	2	3	19	980
Number of fours	1	7	25	1041
Number of fives	1	5	12	975
Number of sixes	1	3	20	978
Expected frequency	1	4	20	1000

We could also use the known (exact) probability to calculate:

$$\text{expected frequency} = \text{probability} \times (\text{number of trials})$$

You can see that the random variations get smoothed out as we do more and more trials, giving a better estimate of the probability.

As we increase the number of throws, the fraction giving a 6 would get closer and closer to $\frac{1}{6}$.

e.g. table above suggests $P(6) = \frac{978}{6000}$.

Adding probabilities

We can add probabilities for events that can never happen together (posh name: "mutually exclusive")

- If cars can be red, green or blue, $P(R \text{ or } G) = P(R) + P(G)$ (it cannot be both red and green!). Can use special symbol,

$$P(R \cup G) = P(R) + P(G)$$

- Raffle tickets (20 numbers) all numbers are equally likely, so the probability of each number being picked is $\frac{1}{20}$. The probability of a 1 or 2 or 3 or 4 or 5 is then

$$\frac{1}{20} + \frac{1}{20} + \frac{1}{20} + \frac{1}{20} + \frac{1}{20} = \frac{5}{20} = \frac{1}{4}$$

Multiplying probabilities

If two different things A and B can happen together, and they do not influence each other (they are "independent"), the probability that they do both happen together is found by multiplying their individual probabilities.

$$P(A \text{ and } B) = P(A) \times P(B)$$

This can be written using a special symbol

$$P(A \cap B) = P(A) \times P(B)$$



"AND" meaning "both together"

e.g. In any month, the probability that I will see a Barn Owl is 0.1

The probability that I will see a Vulcan bomber is 0.5

What is the probability that I see both next month?

- These events are independent so I can multiply the probabilities.
- $P(\text{Barn Owl and Vulcan bomber}) = 0.1 \times 0.5 = 0.05$

Even if the events do affect each other, we can still multiply *provided* we use the correct probability at the instant the event happens (a "conditional probability").

A bag contains 4 white chocolates and 5 dark chocolates. I eat one at random, then another.

$$P(\text{White then Dark}) = \frac{4}{9} \times \frac{5}{8} = \frac{5}{18} \quad (\text{when I pick the second one, there are only 8 chocolates left})$$

Summary

- Relative frequency is an estimate of the probability. More trials give a better estimate.
- Multiply probability by number of experiments to estimate likely number of successes
- Subtract from 1 to find the only unknown probability
- Add probabilities of A and B (if cannot happen together) to get $P(\text{both})$.
- Multiply to get $P(A \text{ and } B \text{ together})$.

Combinations of events: making a list of equally likely outcomes

For combinations of two sets of items, numbers, results etc we can make a table.

eg (a) A box contains 2 milk chocolates (let's call them M1, M2) and a Dark chocolate (D).

I take out a chocolate at random, then replace it, then again take one at random.

What is the probability I get a milk chocolate both times?

There are 9 equally likely options, so the probability of each is $\frac{1}{9}$.

(M1, M1)	(M1, M2)	(M1, D)
(M2, M1)	(M2, M2)	(M2, D)
(D, M1)	(D, M2)	(D, D)

4 of these are "milk then milk", so $P(\text{milk then milk}) = \frac{4}{9}$.

(b) If I did not replace the first chocolate the combinations (M1, M1), (M2, M2) and (D,D)

would be impossible, giving $P(\text{milk then milk}) = \frac{2}{6} = \frac{1}{3}$

	(M1, M2)	(M1, D)
(M2, M1)		(M2, D)
(D, M1)	(D, M2)	