We often want to know likelihood of an event occurring.
Quantify this with dimensionless parameter: Probability (0→1)
Important concept in e.g. quantum physics
Often use histograms to visualise this

x is divided into “bins” of width $\Delta x$

y axis (ordinate) shows number of times your data takes a value from $x \rightarrow x + \Delta x$

Each x bin i contains a number $N_i$
In fact it is $N_i$ divided by bin width $\Delta x$

Thus histogram becomes a probability density since

$$\sum_i \frac{N_i}{N_{tot}} \times \Delta x = \sum_i \frac{N_i}{N_{tot}} = 1$$
Thus a normalised histogram is a probability density!
Consider die throwing experiment. Throw die 6 times and histogram the results.

\[ \Delta x = 1 \text{ so } N_i / \Delta x = N_i \]

\[ N_{\text{tot}} = \sum_i N_i = 6 \]

Total Probability = \[ \sum_i \frac{N_i / N_{\text{tot}}}{\Delta x} \times \Delta x = \sum_i \frac{N_i / 6}{1} = 1 \]

Shape remains the same
Scale changes by constant factor

Now let \( \Delta x \to 0 \)
Histogram \( \to \) continuous distribution (for non-discrete observables)
compare dice throwing result to prob of bus arriving in time \( \Delta t \)
Let's halve the bin width:

\[
\text{Bin content} = \frac{1}{2} \frac{N_i / N_{\text{tot}}}{\Delta x} \approx \frac{N_i}{N_{\text{tot}}} \times \Delta x
\]
i.e. it is unchanged!

Given enough data:
\( \Delta x \to dx \)
\( N_i / \Delta x \to f(x) \)

Normalisation condition is then
\[ \int_{-\infty}^{\infty} f(x) dx = 1 \]

Prob of bus arriving in time \( dt = 1 \text{ns} \) is tiny, but summed over \( 10^{11} \) 'bins' (~1 min) prob becomes sensible value.
Binomial Distribution

- Let's look at the discrete case of a dice throwing experiment.
- Calculate probability of different outcomes.
- Result is a discrete variable.
  - A die can only result in 1, 2, 3...6 i.e. not 2.45.
  - A coin toss only results in heads or tails (never h and t!)
- Consider 6 tosses of a coin (equivalent to single toss of 6 coins).
- How often does this result in 6 heads (6h)?
  - \( P(6h) = \left(\frac{1}{2}\right)^6 = \frac{1}{64} = 1.6\% \)
- Repeat this experiment 100 times:
  - We expect 6h to happen 1 or 2 times (maybe 0 or 3 times).

- What is prob of 1 head & 5 tails i.e. \( P(1h) \)?
  - \( \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^5 \) ? NO! This is \( P(httttt) \) what about \( P(httttt) \) etc?
  - \( P(1h) = 6 \left(\frac{1}{2}\right)^1 \times \left(\frac{1}{2}\right)^5 \) allows h on any of the 6 tosses.
  - \( P(1h) = 6/64 = 9.4\% \) Quite likely to occur in 100 identical experiments.

- What about \( P(4h) \)?
  \( P(4h) = \left(\frac{1}{2}\right)^4 \times \left(\frac{1}{2}\right)^2 \times N \)
  \( N = \) Number of combinations of getting 4 out of 6.

\[
\begin{array}{c|c|c}
\text{No. of h} & N & \text{Examples} \\
\hline
0 & 1 & h h h h t t \\
1 & 6 & h h h t h t \\
2 & 15 & h h t h h t \\
3 & 20 & \\
4 & 15 & h h t t h t \\
5 & 6 & h t h h t t \\
6 & 1 & h t h t h t \end{array}
\]

4 h is like 2 h etc... 

Easier way to determine \( N \): \( \binom{n}{k} = \) number of combinations of \( k \) from \( n \)

\[
\binom{n}{k} = \frac{n!}{(n-k)!k!}
\]
Binomial Distribution

Number of ways of picking 4 heads from 6 tosses is:

\[ _6C_4 = \frac{6!}{(6-4)!4!} = \frac{6 \times 5}{2} = 15 \]

\[ P(4h) = _6C_4 \left( \frac{1}{2} \right)^4 \left( \frac{1}{2} \right)^2 = \frac{15}{64} = 0.234 \quad (= 23.4\%) \]

We took \( P(h) = P(t) = \frac{1}{2} \)

i.e. \( P(\text{success}) = P(\text{fail}) \)

What if \( P(\text{success}) \neq P(\text{fail}) \)?

Let \( p = P(\text{success}) \)

\[ f(k; n, p) = \frac{n!}{(n-k)!k!} p^k (1-p)^{n-k} \]

\[ = _nC_k p^k (1-p)^{n-k} \]

\[ = _nC_k p^k q^{n-k} \quad \text{where} \quad q = 1 - p \quad q=P(\text{fail}) \]

Binomial distribution for various \( n \) values with \( p=0.4 \)

Note the distribution is discrete - lines between points are to guide the eye

\[ f(k;n,p) \] is asymmetric around the peak

Asymmetry reduces as \( n \) increases
Binomial Distribution

- Binomial distribution describes statistics of situations with a true/false outcome only
- Distributions have some nice properties:

\[ \bar{k} = \text{mean number of successes} = \sum_{k=0}^{n} k \times f(k) = np \]

i.e. mean number of heads from 6 tosses = 3

Standard deviation: \[ \sigma_k = \sqrt{np(1-p)} \]

\[ = \sqrt{npq} \]

In general \( p \neq 0.5 \) e.g. Probability(England winning next match)

Binomial statistics used for true/false type experiments

Example: opinion polls

Ask a sample of people a question. Allow only yes/no answers

"Would you prefer to see the re-introduction of the death penalty to the UK?"

Poll 1000 randomly selected people

- \( n = 1000 \)
- \( \text{Yes} = 500 \quad \text{No (or "Don't know") = 500} \)

Thus \( p = 500/1000 = 50\% \) (!)

What is the uncertainty?

\[ \sigma = \sqrt{np(1-p)} \]

\[ = \sqrt{1000 \times 0.5 \times 0.5} \]

\[ = \sqrt{250} = 16 \]

Fractional error \( \frac{\sigma}{np} = \frac{16}{500} = 3\% \)
• To reduce this error by factor of three to 1% need to work very hard

\[ \frac{\sqrt{n \times 0.50 \times 0.50}}{n \times 0.50} = 1\% = \frac{1}{\sqrt{n}} \]

\[ \sqrt{n} = \frac{1}{0.01} \]

\[ n = 10000 \]

Need to increase \( n \) by factor of 10 for a factor \( \sqrt{10} \) reduction in error!