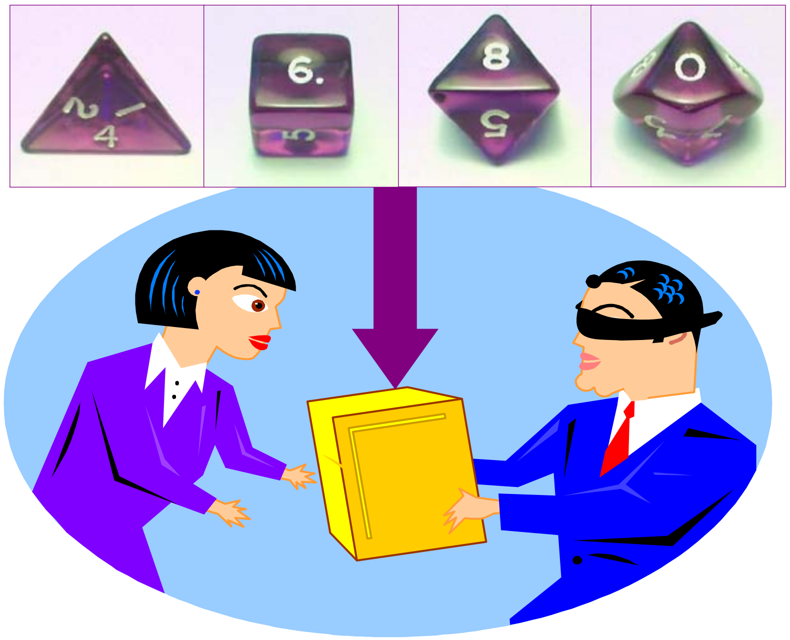
Maximum Likelihood

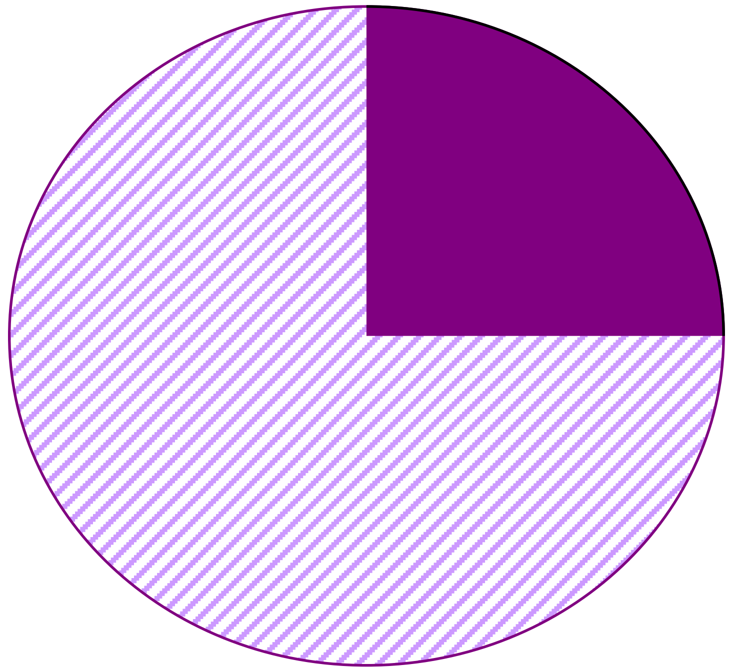
A game of dice



- Put four unbiased dice in a box
- I select a die at random
- How often will you guess correctly which die I selected?

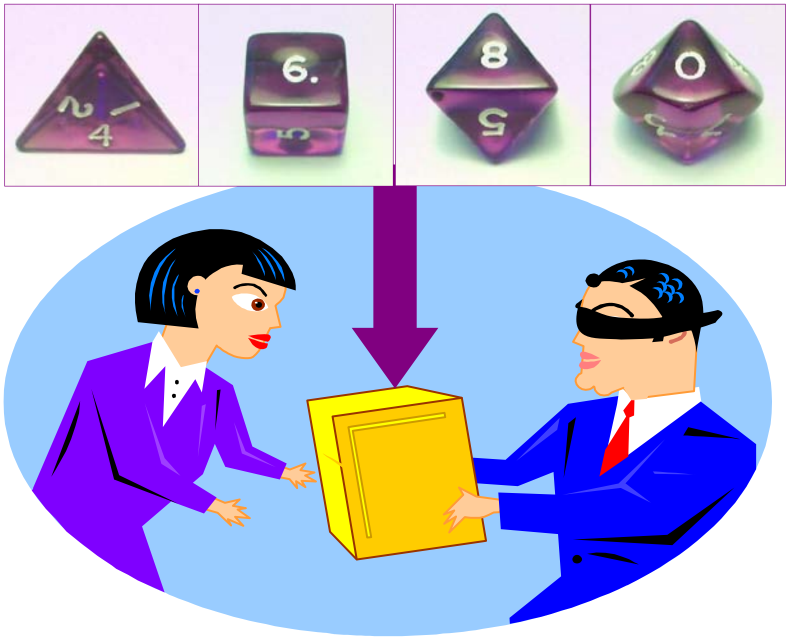
Probability

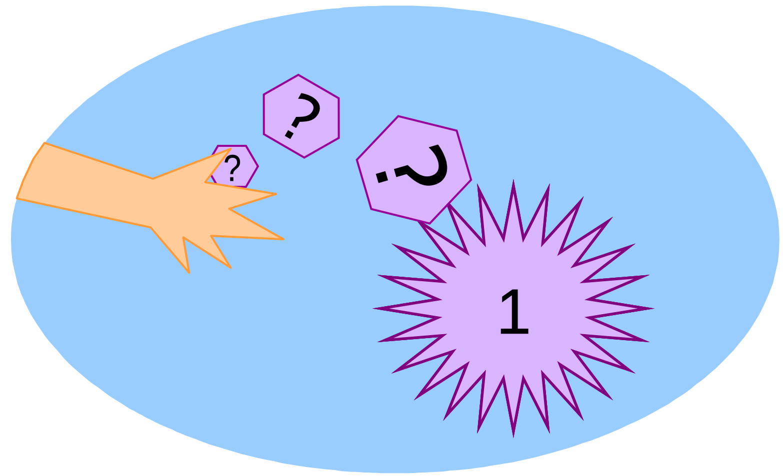
- In the game of dice you have a 1 in 4 chance of being right
- If a large number of people guessed, one quarter would be right each time
- If you play the game many times, you will be right a quarter of the time



Maximum likelihood

A game of dice with data





- Put four unbiased dice in a box
- I select a die at random
- I roll the die and tell you the result of the roll
- Which die did I most likely select?

Maximum likelihood

Roll a 10

- The die obviously must have been the 10 sided die
- What does “must” mean in probabilities?

10

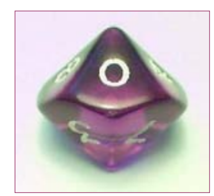
$$P\left(10;\boxed{10}\right) = \tfrac{1}{10}$$

$$P\left(10;\boxed{8}\right) = 0$$

$$P\left(10;\boxed{6}\right) = 0$$

$$P\left(10;\boxed{4}\right) = 0$$

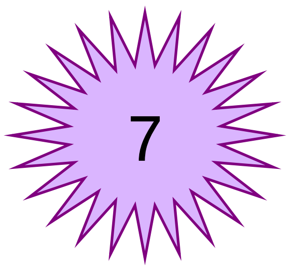
most likely



Maximum likelihood

Roll a 7

- The die could have been the 10 sided or the 8 sided die
- Which die is most likely?



$$P\left(7;\boxed{10}\right) = \tfrac{1}{10}$$

$$P\left(7;\boxed{8}\right) = \tfrac{1}{8}$$

$$P\left(7;\boxed{6}\right) = 0$$

$$P\left(7;\boxed{4}\right) = 0$$

most likely 

Maximum likelihood

Roll a 1

- Could have been rolled by any of the dice
- The most likely die is the one with the highest probability of generating the data

1

$$P\left(1;\boxed{10}\right) = \tfrac{1}{10}$$

$$P\left(1;\boxed{8}\right) = \tfrac{1}{8}$$

$$P\left(1;\boxed{6}\right) = \tfrac{1}{6}$$

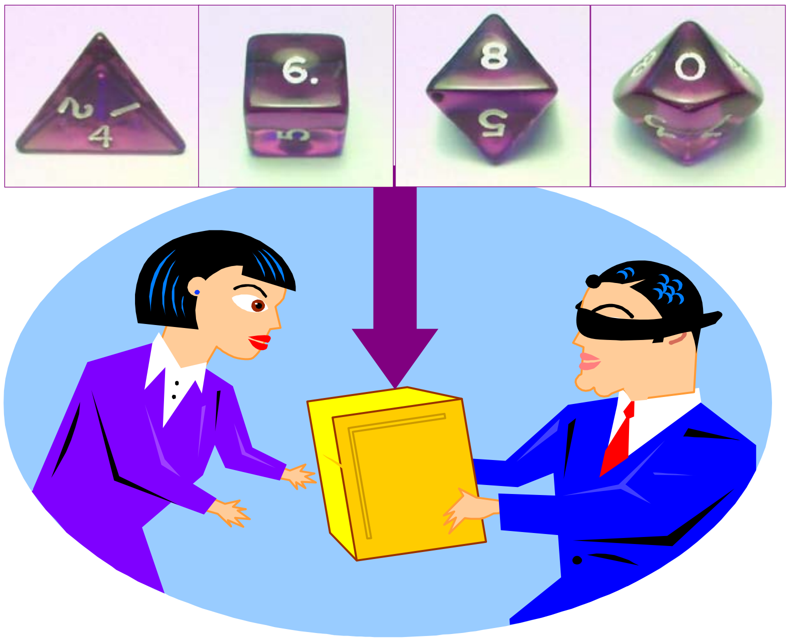
$$P\left(1;\boxed{4}\right) = \tfrac{1}{4}$$

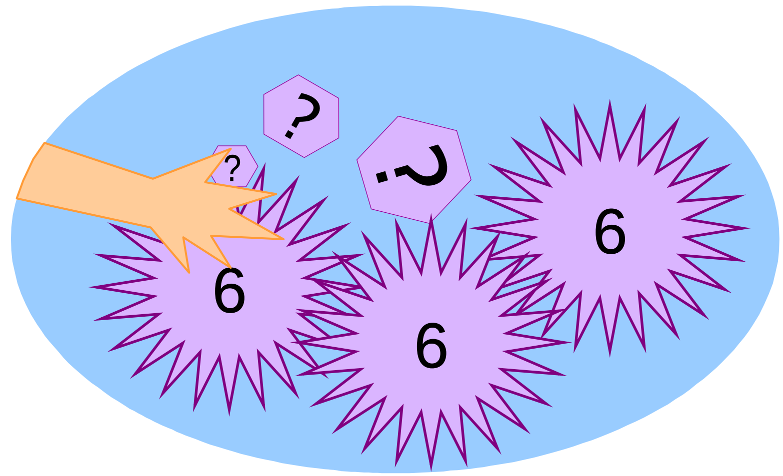
most likely 

Independence

and log-likelihood

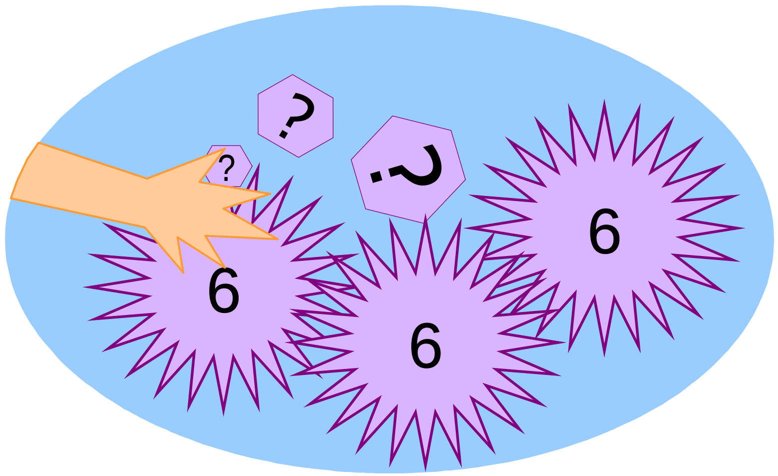
A game of dice with more data





- Put four unbiased dice in a box
- I select a die at random
- I roll that die <u>three</u> times and tell you the results
- Which die did I most likely select?

A game of dice with more data



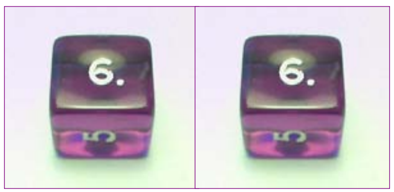
- What is the chance of throwing a 6 three times from a 6-sided die?
- The chance of throwing a 6, or any other number, the second, or third time is not influenced by the value of the first roll - they are <u>independent</u>

Independence and log-likelihood

Multiplying probabilities

- When probabilities are independent they multiply

$$P\left(6;\boxed{6}\right)=\tfrac{1}{6}=0.16666667$$



$$P\left(6,6;\boxed{6}\right)=\tfrac{1}{6}\times\tfrac{1}{6}=\tfrac{1}{36}=0.0277778$$

$$P\left(6,6,6;\boxed{6}\right)=\tfrac{1}{6}\times\tfrac{1}{6}\times\tfrac{1}{6}=\tfrac{1}{216}=0.0046296$$

100 times

$$P\left(6\ldots\times100;\boxed{6}\right)=6^{-100}=1.53064\times10^{-78}$$

Independence and log-likelihood

Computers and small numbers

"Oh great one, what is the probability of throwing a 6 from a six sided die one billion times?"

```
> SYSTEM-F-FLTOVF_F, arithmetic fault,
floating overflow at PC=00006244,
PSL=03C0 0020 %TRACE-F-TRACEBACK,
symbolic stack dump follows
module name       routine name   line
OVERF             OVERF          104
DPARA$MAIN        DPARA$MAIN     276
```

Computers can not store numbers very close to zero

Computers and log(small numbers)

"Oh great one, what is the logarithm of the probability of throwing a 6 from a six sided die one billion times?"

```
> -778151250.4
```

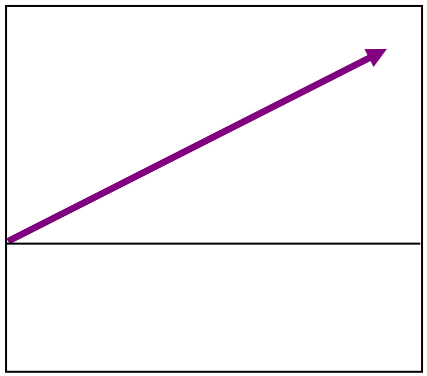
log(likelihood) is not close to zero

- So the log(likelihood) solves the small number problem
- But can we just switch to using the log(likelihood)?

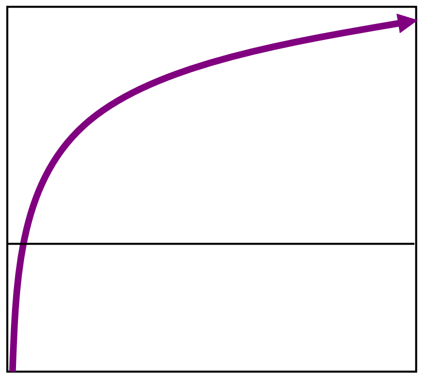
Independence and log-likelihood

Optimisation and logarithms

- Logarithmic functions are "monotonic" functions
  - *i.e.* they "preserve the given order"
  - If $y_1 < y_2$ for all $x_1 < x_2$ then $\log(x_1) < \log(x_2)$
- The parameter values obtained optimising log(likelihood) are the same as those obtained optimising likelihood
  - **Optimising log(likelihood) ≡ Optimising likelihood**



$y = x$



$y = \log(x)$

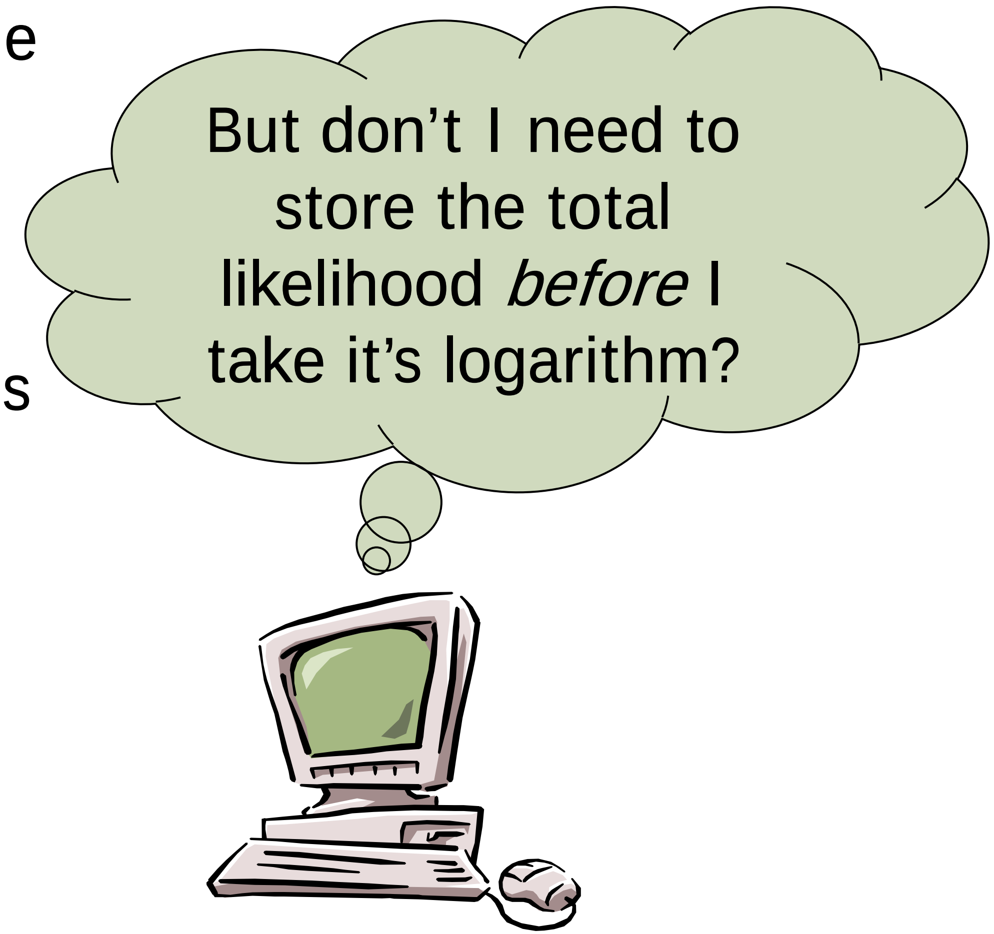
Independence and log-likelihood

Logarithms, products and sums

- No, there is a shortcut to the log(total likelihood) when total likelihood is a product of likelihoods
- If log(total likelihood) equals log(Π likelihoods)

  product
- Then log(total likelihood) also equals Σlog(likelihoods)

  sum



Logarithms and independence

$$\log(\prod \text{likelihoods}) = \sum \log(\text{likelihoods})$$

$$\begin{aligned}
&\log\left(P\left(3,3;\boxed{6}\right)\right)\\
&\quad = \log\left(P\left(3;\boxed{6}\right)\times P\left(3;\boxed{6}\right)\right)\\
&\quad = \log\left(\tfrac{1}{6}\times\tfrac{1}{6}\right)\\
&\quad = \log(0.0277)\\
&\quad = -1.556
\end{aligned}$$

$$\begin{aligned}
&\log\left(P\left(3,3;\boxed{6}\right)\right)\\
&\quad = \log\left(P\left(3;\boxed{6}\right)\right)+\log\left(P\left(3;\boxed{6}\right)\right)\\
&\quad = \log\left(\tfrac{1}{6}\right)+\log\left(\tfrac{1}{6}\right)\\
&\quad = -0.778-0.778\\
&\quad = -1.556
\end{aligned}$$

Independence and log-likelihood

Minimising

- Computer algorithms are designed to minimise
- Therefore we optimise our parameters by minimising the -log(likelihood)