Data, Probability and Statistics

Janine McIntosh: janine@amsi.org.au

Michael O'Connor: moconnor@amsi.org.au

Probability

Have you all chosen your biscuit?

If not, please do so now.

Probability

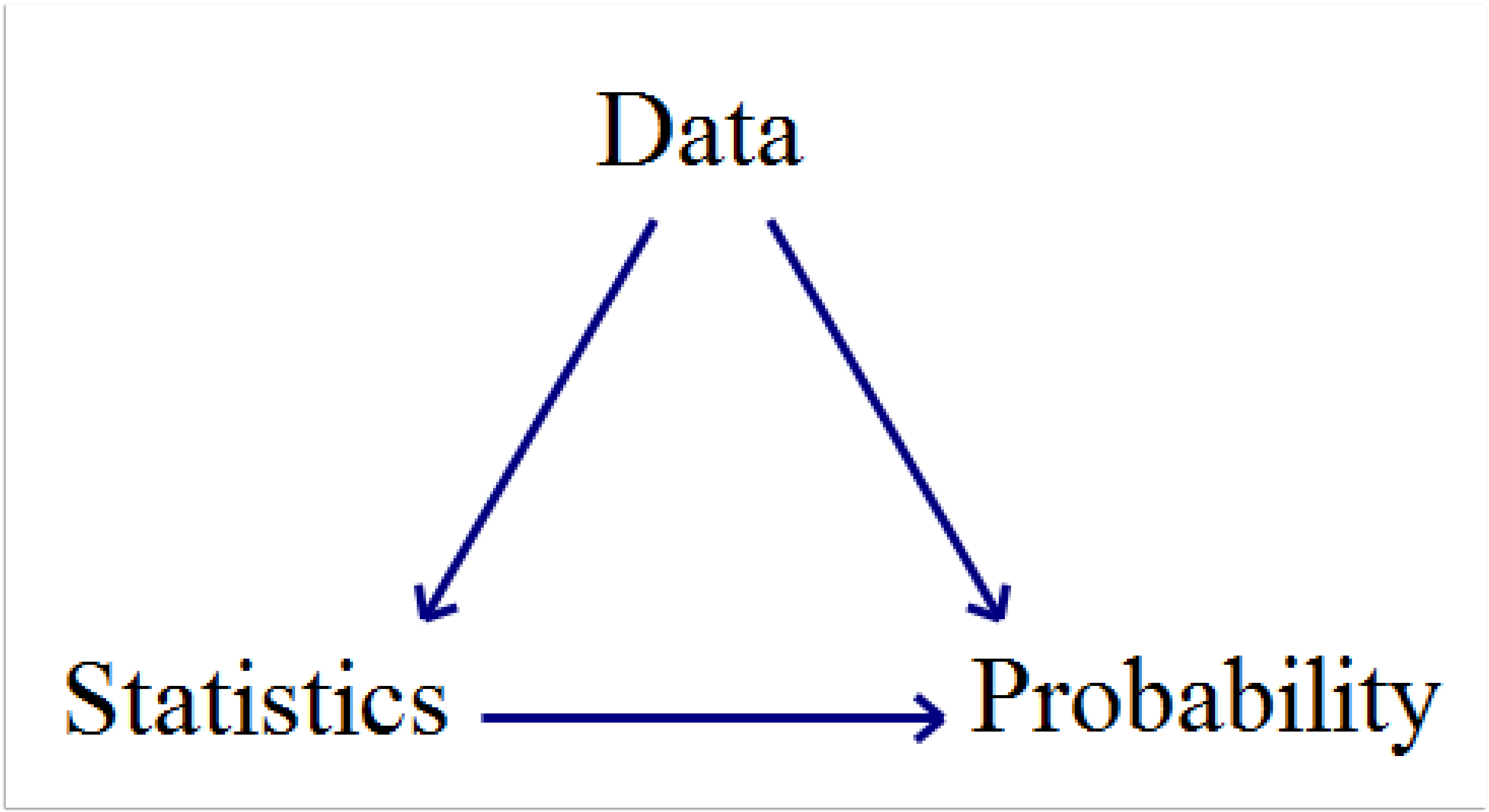
What is it?

Chance – a layman's expression for the possibility of something happening.

Probability – The calculation of the precise likelihood of something happening.

Probability

What is it?



Probability

What is it?

Data – Information about properties and values associated with an item of interest

Statistics – calculations and graphical representations to make sense of the raw information

Probability – Extrapolation of the data and statistics to make predictions about how likely different outcomes are to occur.

Probability

What is it?

Some probabilities are fixed. That is to say, when playing a game with a six sided die, the probability of rolling a six on any one throw is always $\frac{1}{6}$.

Other probabilities change with time and circumstance: for example, forty years ago males who smoked was 45% of the population while in 2010 it was just under 18% (17.9%)

Probability

Experiment vs Theory?

Experiment –

Theory -

Probability

Experiment vs Theory?

Which of the four rows is not (least likely to be) random?

T	T	H	T	H	T	T	T	H	T	T	T	T	T	T	H	T	T	T	T	T	T	T	H	H	H	T	T	T	T
T	T	T	T	T	T	T	H	T	T	H	T	T	H	T	T	T	T	H	T	T	H	T	T	H	T	T	T	H	H
H	H	T	H	T	H	T	T	H	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H	T	H
H	H	H	T	T	H	H	H	H	H	T	H	H	H	H	T	H	T	H	H	H	H	H	H	H	H	H	H	H	H

Probability

Why is it important?

Table 1.3.2

Percentage of current smokers* in Australia, 1945–1976

Year	Male (%)	Female (%)
1945	72	26
1964	58	28
1969	45	28
1974	45	30
1976	43	33

[Tobacco in Australia: Facts and Issues (Cancer Council)](http://www.tobaccoinaustralia.org.au/downloads/chapters/Ch1_Prevalence.pdf)

http://www.tobaccoinaustralia.org.au/downloads/chapters/Ch1_Prevalence.pdf

Probability

Why is it important?

Angelina Jolie

BRCA 1 gene from her mother

87% probability of developing breast cancer

After surgery: 5% probability

Probability

Why is it important?

Medical Research

Insurance

Games of Chance

+ + +

Probability

WHY DO WE NEED IT?

Probability is an area of mathematics with many diverse applications.

For example, it is used in measuring risk in insurance and banking and to set both interest rates and insurance premiums.

Probability calculations are an essential component of most medical research. Research in genetics often involves complicated probability considerations.

It is also closely linked to mathematical and applied statistics.

Probability

WHY DO WE NEED IT?

Statistics and statistical thinking have become increasingly important in a society that relies more and more on information and calls for evidence.

The need to develop statistical skills and thinking across all levels of education has grown and is of core importance in a century which will place even greater demands on society for statistical capabilities throughout industry, government and education.

Probability

WHY DO WE NEED IT?

Statistics is the science of variation and uncertainty.

Concepts of probability underpin all of statistics, from handling and exploring data to the most complex and sophisticated models of processes that involve randomness.

Statistical methods for analysing data are used to evaluate information in situations involving variation and uncertainty, and probability plays a key role in that process.

Probability

WHY DO WE NEED IT?

All statistical models of real data and real situations are based on probability models.

Probability models are at the heart of statistical inference, in which we use data to draw conclusions about a general situation or population of which the data can be considered randomly representative.

Probability

How do we teach it?

Biscuits and Birthdays

What biscuit was most popular?

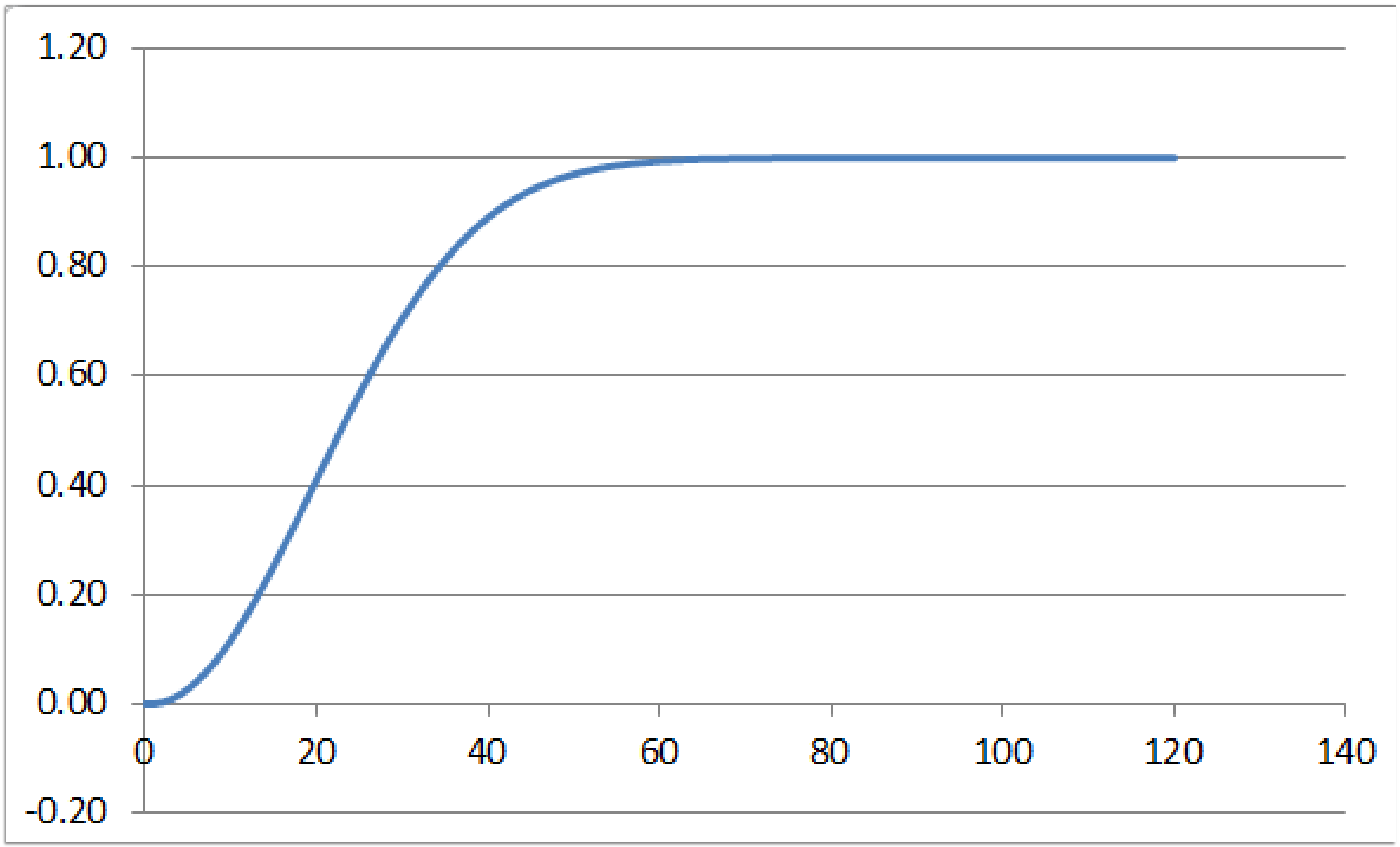
When attending an AMSI Cluster PD, what is the probability of a random participant having a Turkish Delight Tim Tam?

What is the probability that two people in this room share the same birthday?

Probability

How do we teach it?

Sharing a Birthday



$$P(n) = 1 - \frac{365!}{(365-n)! \times 365^n}$$

Probability

How do we teach it?

Not all teaching can be so immediate or tasty.

ACCEPT that not everything can be contextualised,

BUT give them glimpses of how it will affect their lives now AND IN the future.

Probability

Applications

It is worth having the conversation with students that until very recently probability and statistics have been interesting...

...but a not very useful or accessible to most people.

Probability

Applications

robbelles.wordpress.com

It is only now, with

- the power of modern computers and
- the ability to collect enormous amounts of data

that the power of probability is becoming evident.

Probability

Applications

In the 21st Century we use statistical analysis to

- understand incredibly complex structures and patterns (bioinformatics, wine)

- make more accurate decisions (stockmarket, sports statistics)

robbelies.wordpress.com

Probability

Applications

Every industry can now take advantage of this:

Some examples include:

Biology,	Business,	Computing,
Criminology,	Defence,	Energy,
Environment,	Epidemiology,	Finance,
Government,	Health,	Insurance,
Market Research,	Meteorology,	Operations Research,
Transport	Teaching.	

Probability

Applications

Now, for perhaps the first time in history,

Laypeople can name an industry and ask,

“Where can we use mathematics?”

instead of trying to think of industry situations where is mathematics is useful.

Every decision we can make in relation to data, we can now pose as a mathematical problem to be solved.

Probability

Applications

We are used to checking the weather forecast at the end of the TV news bulletin every evening.

On most occasions the presenter will confidently state “It will rain heavily tomorrow” or “A cool change will arrive at about 10pm tonight”.

Probability

How do we teach it?

AND NOW FOR THE CONTENT

Probability

How do we teach it?

Probability is about counting

1) How many possible outcomes an event has

AND

2) How many of those outcomes are desirable

How many biscuits are there and how many do I like?

Probability

The Language of Probability

In order to count effectively we need an appropriate vocabulary. This includes:

experiment

sample space

outcome

equally likely outcome

probability

total probability

subset

complement

event

Probability

The Language of Probability

Experiment

for example:

rolling a die

flipping a coin

drawing a card from the deck

choosing a biscuit

Probability

The Language of Probability

Outcome

the possible results
in the experiment:

heads or tails

1,2,3,4,5 or 6

Probability

The Language of Probability

sample space

the complete set of possible outcomes:

$S = \{1,2,3,4,5,6\}$

Probability

The Language of Probability

equally likely outcome

when no outcome should occur more frequently than another

Probability

The Language of Probability

probability

the chance, possibility or likelihood of a particular outcome occurring

Probability

The Language of Probability

total probability

there are a finite number of possible outcomes, so the sum of the probability of these outcomes is 1.

Probability

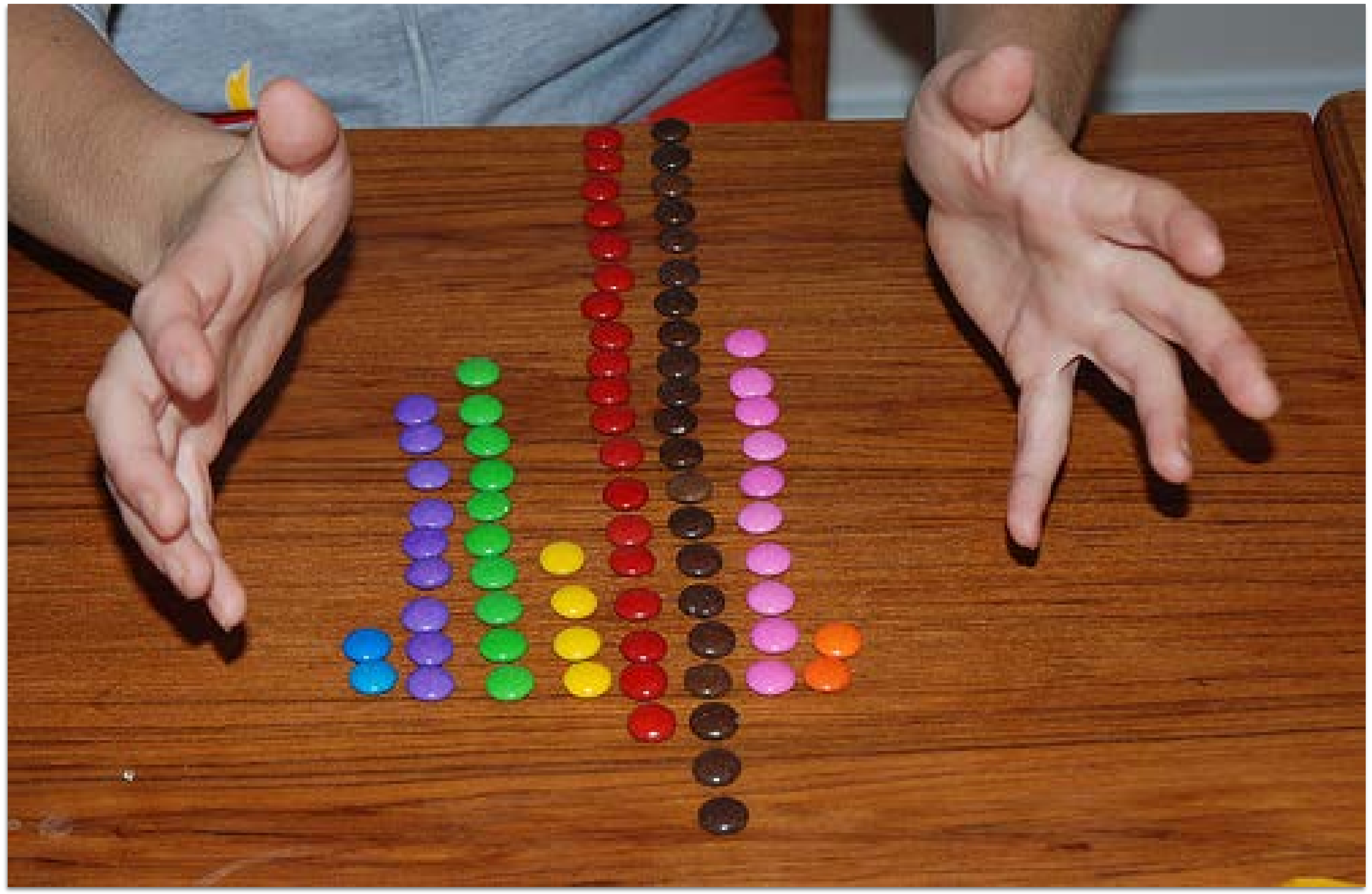
The Language of Probability

subset

a set containing some, all or none of the elements of the sample space e.g.

red smarties

A



Probability

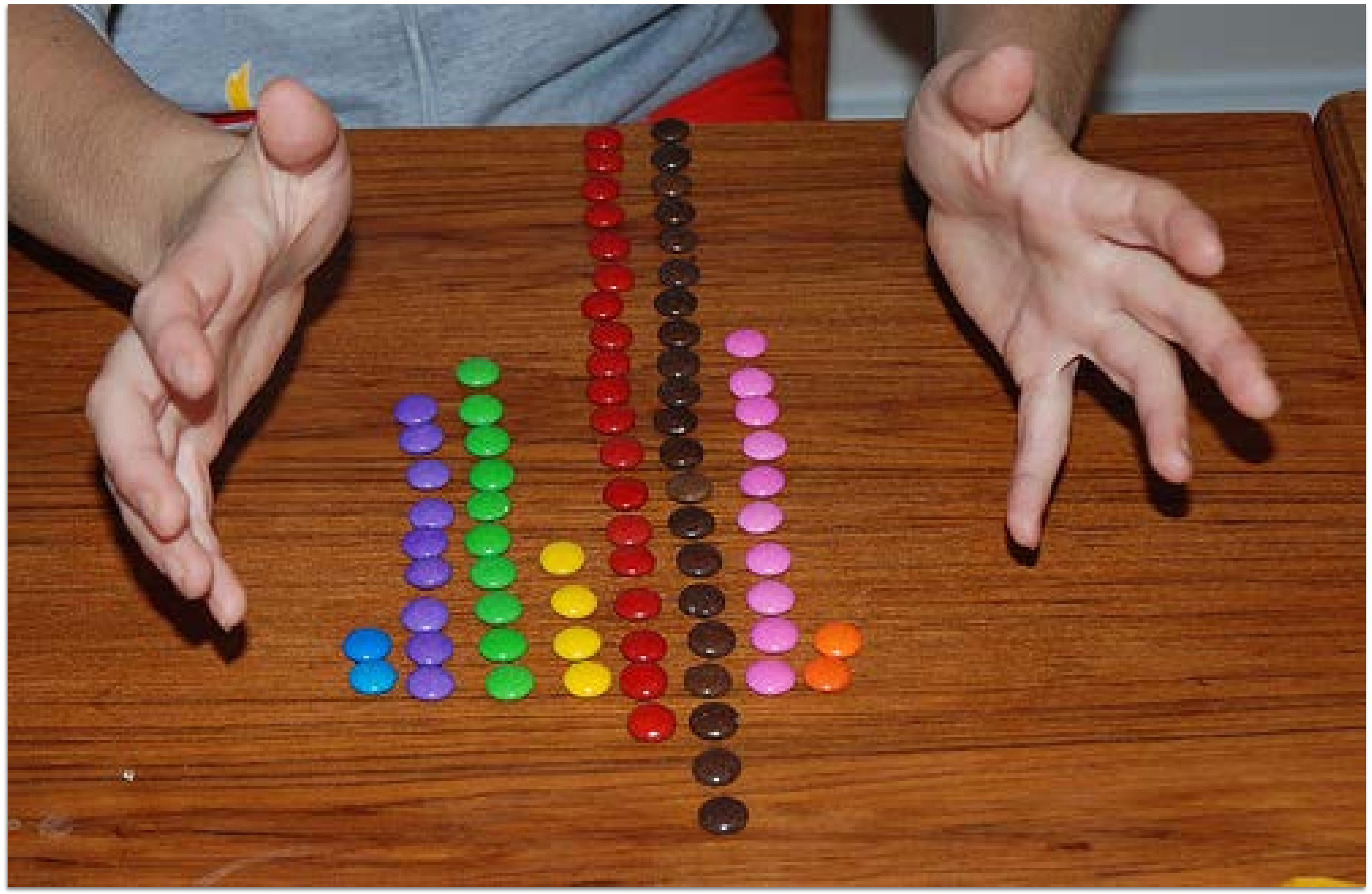
The Language of Probability

complement

all the other elements in the sample space that are not in the in a subset *A e.g.*

all other smarties

Ac



Probability

The Language of Probability

event

a subset of the sample space e.g.

an even number
two heads
a red smartie
3 scissors

Probability

Expressing the probability of an event

$$\text{Probability of an event} = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$$

Probability

Expressing the probability of an event

Experiment - A die is tossed once to obtain an even number

Probability P

The event that we are looking for is shown

in brackets next to the P

e.g. P(even)

There are 3 favourable outcomes i.e. 2,4 and 6

There are a total of 6 possible outcomes i.e. 1,2,3,4,5 and 6

Probability

Expressing the probability of an event

$$\text{Probability of an event} = \frac{\text{number of favourable outcomes}}{\text{total number of outcomes}}$$

$$P(\text{even}) = \frac{3}{6}$$

$$= \frac{1}{2}$$

Probability

The Australian Curriculum

Year	Content Description
1	Identifies outcomes of familiar events involving chance and describes them using everyday language such as ‘will happen’, ‘won’t happen’ or ‘might happen’
2	Identifies practical activities and everyday events that involve chance. Describes outcomes as ‘likely’ or ‘unlikely’ and identify some events as ‘certain’ or ‘impossible’

Probability

The Australian Curriculum

Year	Content Description
3	Conducts chance experiments (tossing and coin or drawing a ball from a bag), identifies and describes possible outcomes and recognise variation in results
4	Describes possible everyday events and order their chances of occurring
	Identifies everyday events where one cannot happen if the other happens (eg: weather, if it is dry, it cannot be wet)
	Identifies events where the chance of one will not be affected by the occurrence of the other (new baby is either a boy or a girl does not depend on the sex of the previous baby) (Independence)

Probability

The Australian Curriculum

Year	Content Description
5	Lists outcomes of chance experiments involving equally likely outcomes and represents probabilities of those outcomes using fractions
	Recognises that probabilities range from 0 to 1
6	Describe probabilities using fractions, decimals and percentages
	Conduct repeated trials of chance experiments, identifying the variation between trials and realising the results tend to the prediction with larger numbers of trials
	Compare observed frequencies across experiments with expected frequencies and predict likely outcomes from a run of chance events and distinguish these from surprising results

Probability

The Australian Curriculum

Year	Content Description
7	Construct sample spaces for single-step experiments with equally likely outcomes. Discuss the meaning of probability terminology (eg: probability, sample space, favourable outcomes, trial)
	Assign probabilities to the outcomes of events and determine probabilities for events. Express probabilities in common and decimal fractional and percentage forms

Probability

The Australian Curriculum

Year	Content Description
8	Identify complementary events and use the sum of probabilities to solve problems
	Describe events using language of 'at least', exclusive 'or' (A or B), inclusive 'or' (A or B or both) and 'and'
	Represent such events in two-way tables and Venn diagrams and solve related problems

Probability

The Australian Curriculum

Year	Content Description
9	List all outcomes for two-step chance experiments, both with and without replacement using tree diagrams or arrays. Assign probabilities to outcomes and determine probabilities for events
	Calculate relative frequencies from given or collected data to estimate probabilities of events involving 'and' or 'or'
	Investigate reports of surveys in digital media and elsewhere for information on how data were obtained to estimate population means and medians

Probability

The Australian Curriculum

Year	Content Description
10	Describe the results of two- and three-step chance experiments, both with and without replacements, assign probabilities to outcomes and determine probabilities of events. Investigate the concept of independence. Recognise that some sets of chance events are dependent on a previous result and others are not, that this distinction is important when calculating probabilities, and that events are independent if P(A) x P(B) = P(A and B)
	Use the language of 'if ....then, 'given', 'of', 'knowing that' to investigate conditional statements and identify common mistakes in interpreting such language

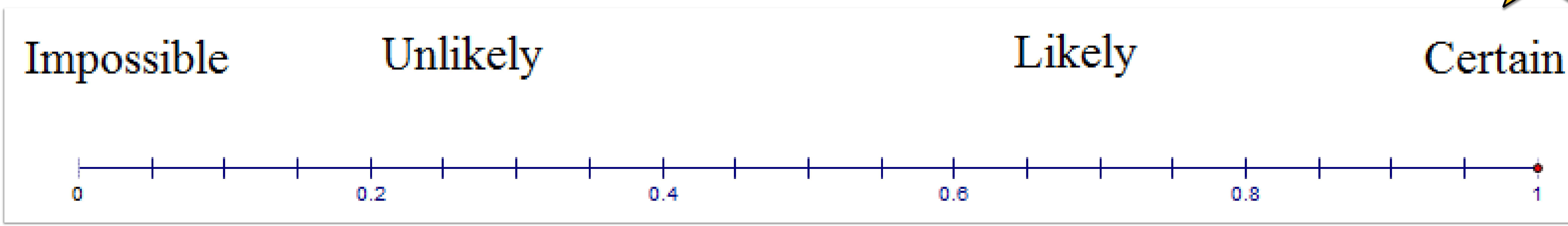
Probability

The Australian Curriculum

Year	Content Description
10 A	Investigate reports of studies in digital media and elsewhere for information on the planning and implementation of such studies, and the reporting of variability. Evaluate media reports that refer to data from a range of contexts.
	Evaluate whether graphs in a report could mislead, and whether graphs and numerical information support the claims

Probability

Expressing the probability of an event



Words	It is likely that it will rain tomorrow.
Fractions	There is a 3/4 chance that it will rain tomorrow
Percentage	There is a 75% chance it will rain tomorrow.
Decimals	There is a 0.75 chance it will rain tomorrow.

Probability

Dice Games

Greedy Pig

First Down the Mountain

Cat and Mouse

Mebourne Cup

Don’t forget the Maths!

Introduction to Probability

Greedy Pig

Equipment: an ordinary 6-sided die

Each turn of the game consists of one or more rolls of the die. You keep rolling until you decide to stop, or until you roll a 1. You may choose to stop at any time.

If you roll a 1, your score for that turn is 0.

If you choose to stop rolling before you roll a 1, your score is the sum of all the numbers you rolled on that turn.

The player with the highest score wins.

Each player has 10 turns.

Introduction to Probability

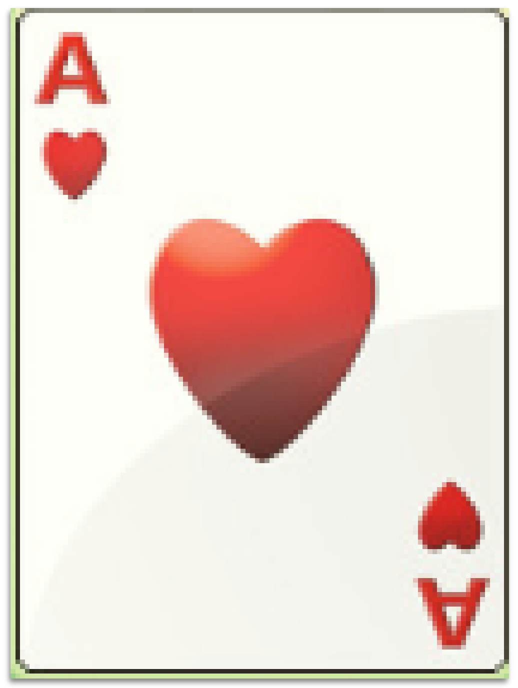
First Down the Mountain

Available from

http://www.blackdouglas.com.au/taskcentre/087first.htm

Probability

Cards



Choose a team

Probability

Cards

Standard packs have 52 cards of 4 suits and 13 cards in each suit.

Many students these days do not know this.

Repetitions and familiarisation. 5 to 10 minutes regularly

Use card games and card tricks “Pick a card, any card”

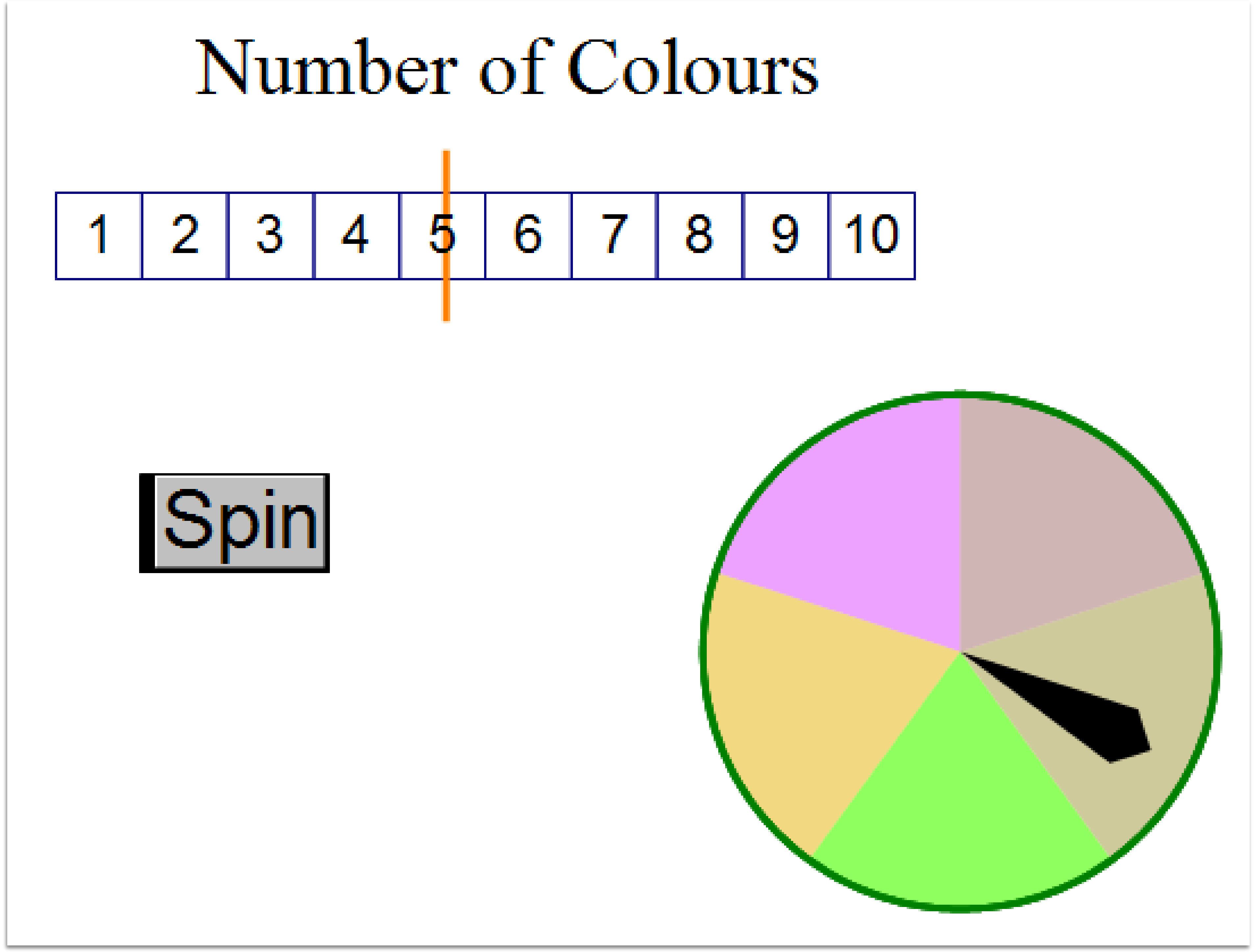
Use suits (Hearts, Diamonds, Clubs, Spades) to choose teams. Count how many in each team.

Probability

Spinners

Spinners can be physical or digital

They are useful for probabilities where you want the number of outcomes to be non-standard.



Probability

Bag of Cubes

	Red	Blue	Green	Yellow	Orange
Bag 1	1	4	2	2	6
Bag 2	3	6	1	1	4
Bag 3	3	3	3	3	3
Bag 4	5	2	2	4	2

Probability

Mr O'Connor's Die of Death!

Used for randomly selecting students to answer questions in class

It started out as a physical die and then turned into a spread sheet

Keep track of how many times each student is chosen in a lesson. Discuss randomness and fairness, etc

Probability

The Italian Restaurant

100 diners were asked if they had ordered garlic bread, pizza or gelati.

11 people ordered all three items.

29 people ordered pizza and gelati but no garlic bread.

16 people ordered pizza only.

There were a total of 60 orders for pizza.

Nobody ordered just garlic bread and gelati.

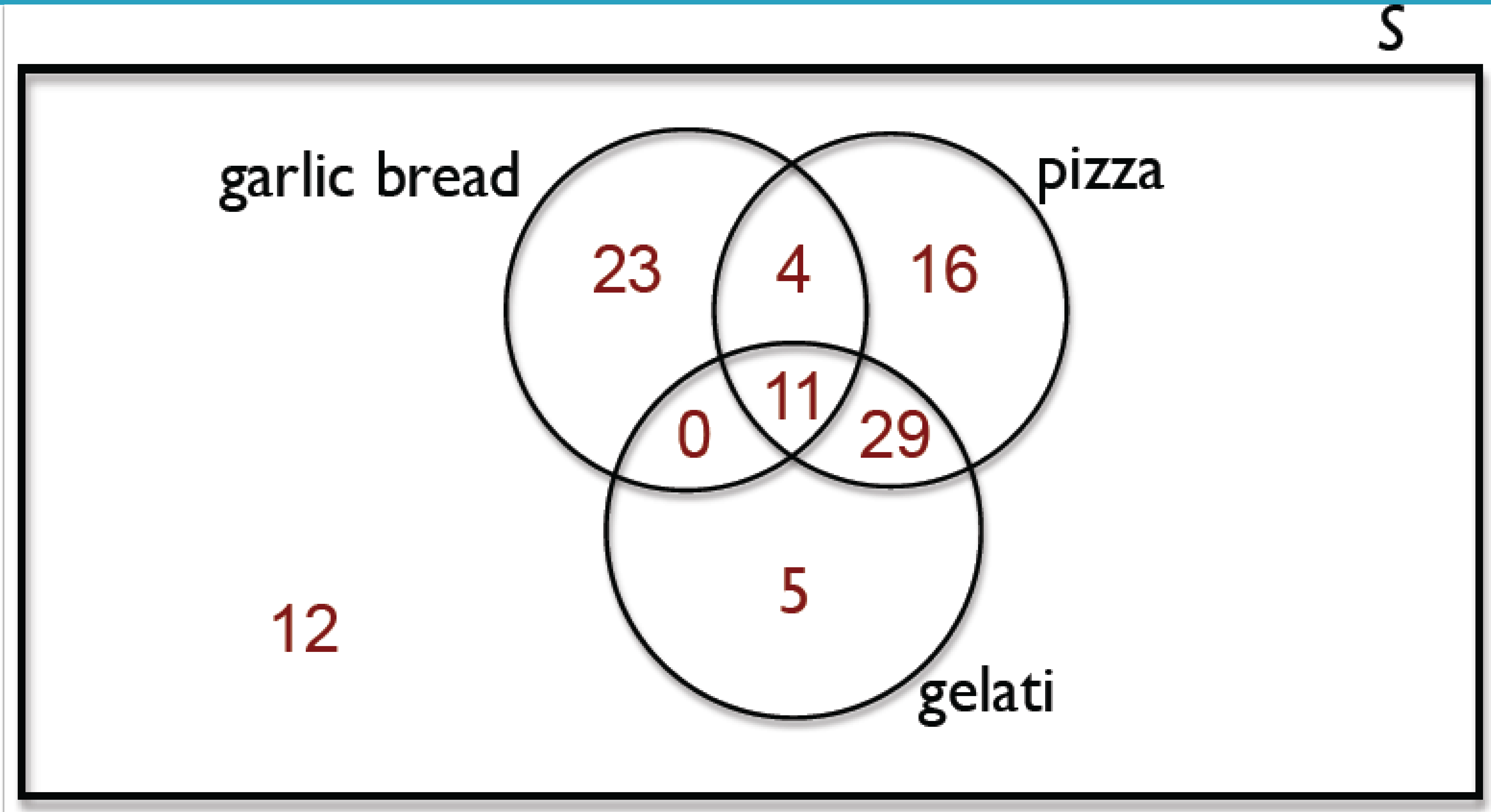
There were a total of 45 gelati orders.

12 people did not order any of these items.

How many people ordered garlic bread only?

Probability

Venn Diagrams



Probability

Resources

Introduction to Probability - TIMES Project , ICE-EM

SAM – Middle Years

maths300 - Education Services Australia

mathsonline -http://www.mathsonline.co.uk/nonmembers/resource/prob/

Probability

Birthdays

We have already looked at sharing birthdays.

What about celebrating birthdays in your class with games that allow for counting and finding probabilities?

Probability

Hands, Arms and Legs

Cross your arms.

Did you go right over left or left over right?

Fold your hands

Which index finger is on top: left or right?

Cross your legs.

Did you go right over left or left over right?

Probability

Rolling two dice and adding

Suppose we have two dice, one red and one blue.

We roll both dice simultaneously and then add their face values.

Probability

Rolling two dice and adding

Do 5 trials.

What did you get?

Probability

Rolling two dice and adding

What are the possible outcomes?

Probability

Rolling two dice and adding

The possibilities can be tabulated:

Probability

Rolling two dice and adding

	+	blue 1	2	3	4	5	6
	1						
	2						
	3						
red	4						
	5						
	6						

Probability

Rolling two dice and adding

		blue					
	+	1	2	3	4	5	6
	1	2	3	4	5	6	7
	2	3	4	5	6	7	8
	3	4	5	6	7	8	9
red	4	5	6	7	8	9	10
	5	6	7	8	9	10	11
	6	7	8	9	10	11	12

Probability

Rolling two dice and adding

There are 36 equally likely rolls but there are only 11 outcomes in our sample space:

2; 3; 4; 5; 6; 7; 8; 9; 10; 11 and 12.

Probability

Misconceptions

Outcomes of an experiment need not be equally likely!

Assuming that each outcome is equally likely is the most common misconception in probability.

Often people say, “Either this will happen or that will happen, so there is a 50% chance that this will happen"; even when the comment is contradicted by their own experience.

Always, Sometimes, Never

https://www.ncetm.org.uk/resources/

Probability

History

In order to agree to a fair contract it is essential to quantify the risk involved.

Shipping insurance contracts date back to Roman times (200 BC – 600 AD). This means that probabilities were being empirically calculated over two thousand years ago. Premiums for ships plying their trade in the eastern Mediterranean could be as high as 50% for a single trip!

Compare this with modern car premiums that range from 2% to 4% for “safe drivers” for a whole year.

Probability

History

Astragals

Sheep knuckles.

Not all sides are equally likely to roll face up

Probability

History

1654 Chevalier de Mere inspired Pascal and Fermat to correspond on games of chance (particularly on how to split the pot of money in a game that has been interrupted)

Probability

History

In the mid-seventeenth century, Pascal and Fermat discussed through correspondence, the basics of probability theory and solved the dividing the wager problem using the binomial theorem, and of course, Pascal's Triangle.

(By the way, Pascal's triangle or the 'arithmetical triangle' was discovered at least six hundred years earlier by an Arabic mathematician.)

Probability

History

The earliest significant record of probabilities being exactly calculated date from the late middle ages in Europe.

Games of chance had been played since the earliest times (at least five thousand years ago) but detailed analysis of, for example, the fifty-six ways three dice could land and their different probabilities was not recorded before about 1400.

Another early discussion concerned the fair (equitable) way to split the stakes in an incomplete game.

Probability

History

Christiaan Huygens

Pierre de Fermat

Probability

History

“It seems that to make a correct conjecture about any event whatever, it is necessary only to calculate exactly the number of possible cases and then to determine how much more likely it is that one case will occur than another. But here at once our main difficulty arises, for this procedure is applicable to only a very few phenomena…. What mortal, I ask, could ascertain the number of diseases, counting all possible cases ... and say how much more likely one disease is to be fatal than another…? Or who could enumerate the countless changes that the atmosphere undergoes every day and from that predict what the weather will be a month or even a year from now?”

Probability

History


JACOBI BERNOULLI,
Profess. Basil. & utriusque Societ. Reg. Scientiar.
Gall. & Pruss. Sodal.
MATHEMATICI CELEBERRIMI,

ARS CONJECTANDI,

OPUS POSTHUMUM.

Accedit

TRACTATUS

DE SERIEBUS INFINITIS,

Et EPISTOLA Gallicé scripta

DE LUDO PILÆ
RETICULARIS.

BASILEÆ,
Impensis THURNISIORUM, Fratrum.
cIↄ Iↄcc xiii.

Probability

History

Pierre de Fermat

Probability

History

The first book on probability was written by Christian Huygens in 1657. Most of his ideas were written in terms of expectation, which he thought of as the average amount that one would win if a game was played many times. It is essentially the average – the average score when a die is thrown is 3 1/2.

The next such book did not appear for over fifty years when James (Jakob) Bernoulli published Ars Conjectandi in 1713, cited earlier.

Probability

History

Christiaan Huygens

Probability

History

Bernoulli introduced a number of novel ideas:

- the law of large numbers and
- the concept of a confidence interval.

http://www.memo.fr/Media/Bernoulli.jpg

Probability

History

Bernoulli, quite arbitrarily, chose a probability of 0.999 to mean that an event was a moral certainty and conversely if the probability of the event was less than 0.001 then the event was morally impossible. As an example, Bernoulli considered an urn containing 3000 white and 2000 black pebbles. By analysing sums of terms in a binomial expansion, he showed that by sampling the urn 25 550 times one would know with moral certainty that the proportion was within 1/50 of the true proportion of 3/5 to 2/5

Clearly this is incredibly sophisticated compared to the elementary counting of Pascal, only fifty years earlier.

Probability

History

Perhaps the most important statistical method of the nineteenth century was that of least squares. This method provided one of the primary tools for what was called “the combination of observations”. A typical problem is to find the line of best fit, $y = a + bx$, given k data points (x_1, y_1), (x_2, y_2) ….. (x_k, y_k).

Many such problems arose in astronomy and Laplace, Legendre and Gauss developed the method of least squares for determining (or estimating) a and b in the line of best fit.

Probability

History

Gauss used the method of least squares to predict the position of the asteroid Ceres from the limited observational data he had available from astronomers.

And it worked!!!!!

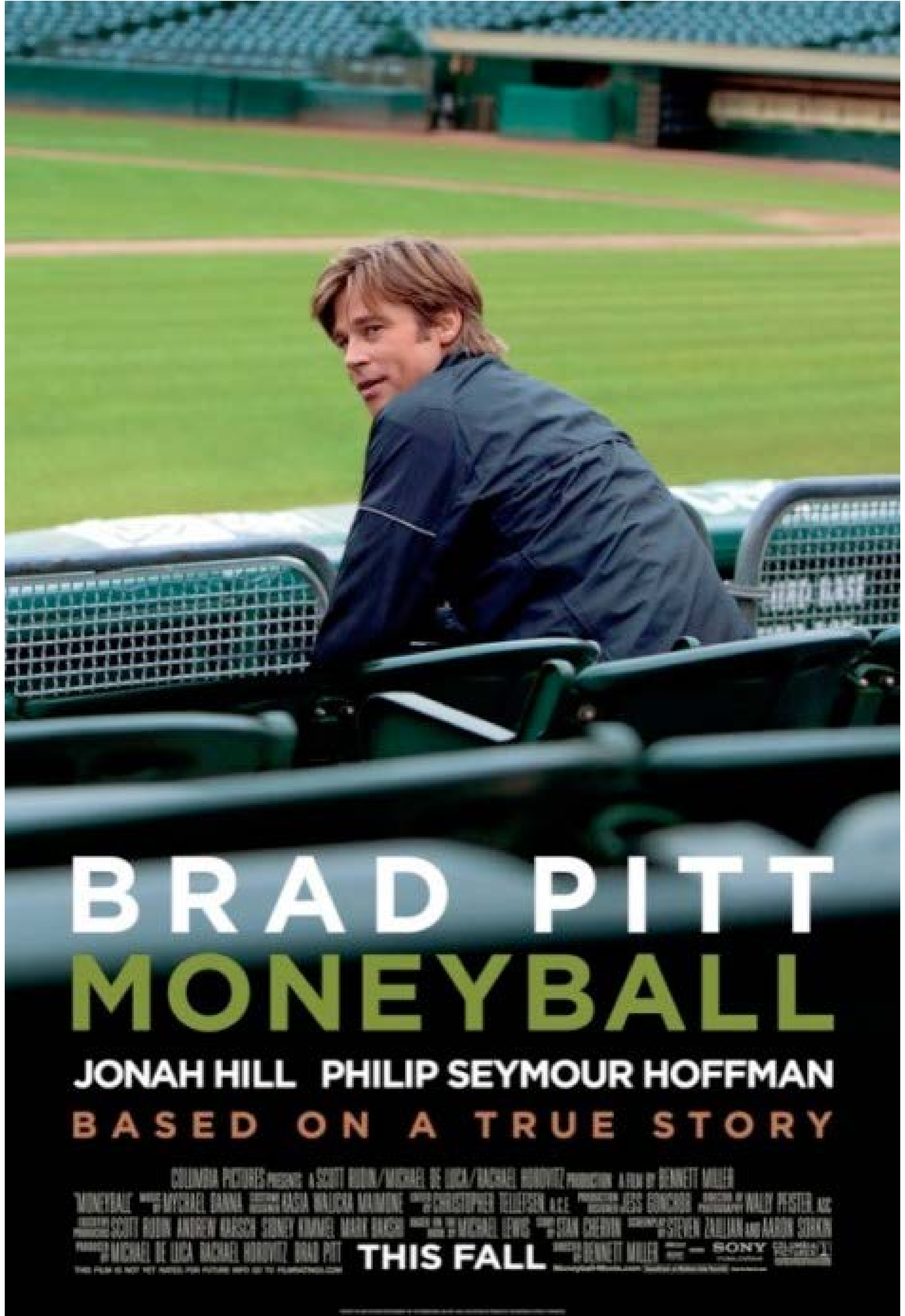
Probability

More Applications - The Value of Wine

Bordeaux wines have been made in much the same way for centuries. The variability in the quality and prices of Bordeaux vintages is predicted by the weather that created the grapes.

The analysis provides a useful basis for assessing market inefficiency, the effect of climate change on the wine industry and the role of expert opinion in determining wine prices.

http://wallpaper.diq.ru/40__Bordeaux_Red_Wine_Glasses.htm



Probability

Applications - Super Bowl

http://blogamole.tr3s.com/2011/09/23/review-show-me-the-moneyball/

Probability

Applications - World Series Baseball

‘Moneyball’ is a movie about Oakland A's general manager Billy Beane's successful attempt to put together a baseball club on a budget by using computer-generated analysis to draft his players.

Billy Beane had the lowest salary constraint in baseball. He found a competitive advantage using probability.

Billy turned baseball on its ear when he used statistical data to analyse and place value on the players he picked for the team.

Probability

Applications – Gambling so as to lose as little as possible

A group of MIT Maths students spend their weekends in Las Vegas

And break the bank

Based on a True story

The DVD has an extra where the actors explain the probability and the mathematics behind the movie.

It also has Monty Hall

Probability

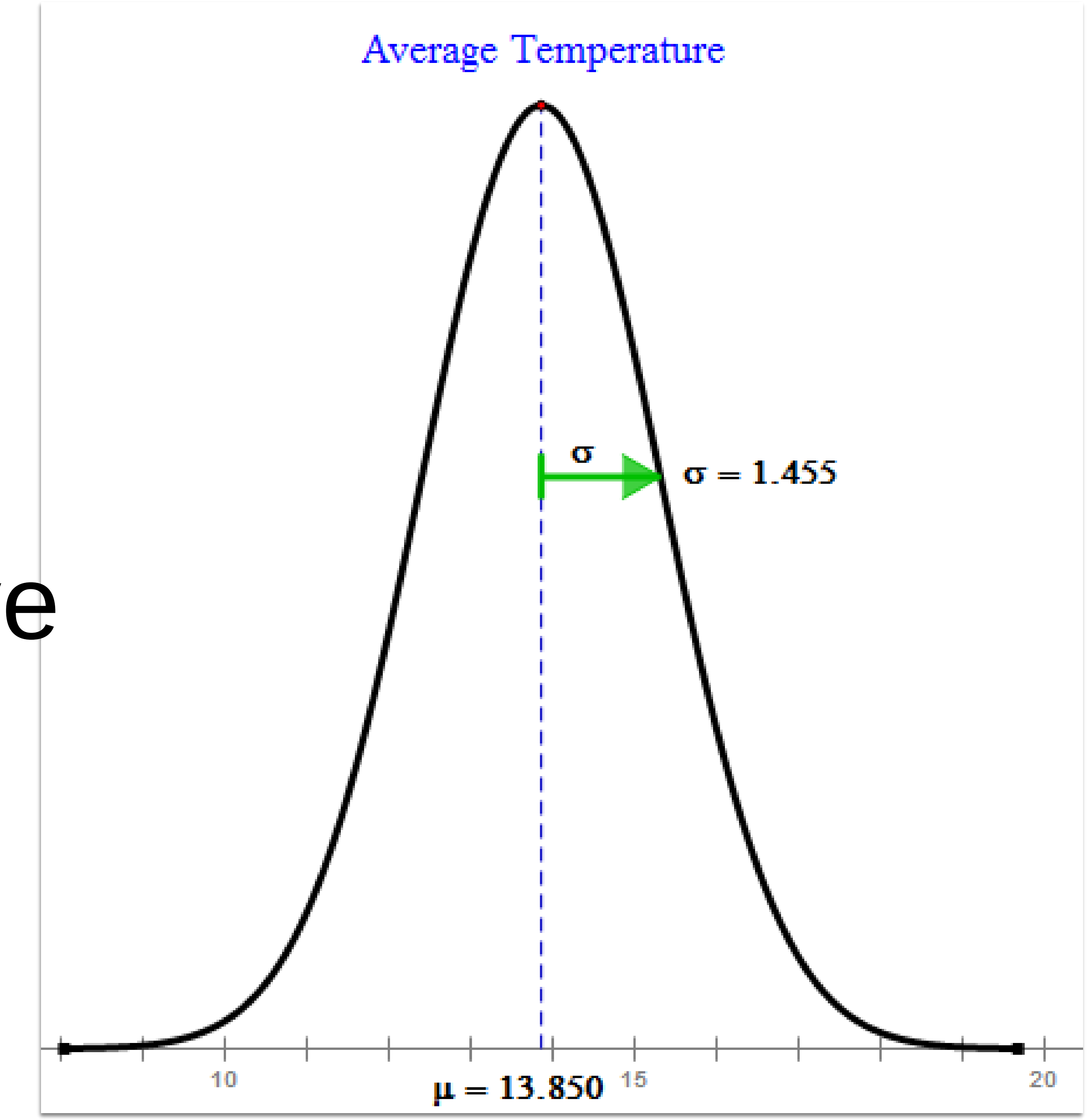
Applications – Explaining Climate Change

What is my one?

The area under the bell curve.

Statistics gives us the shape.

Probability is the area under the curve



Probability

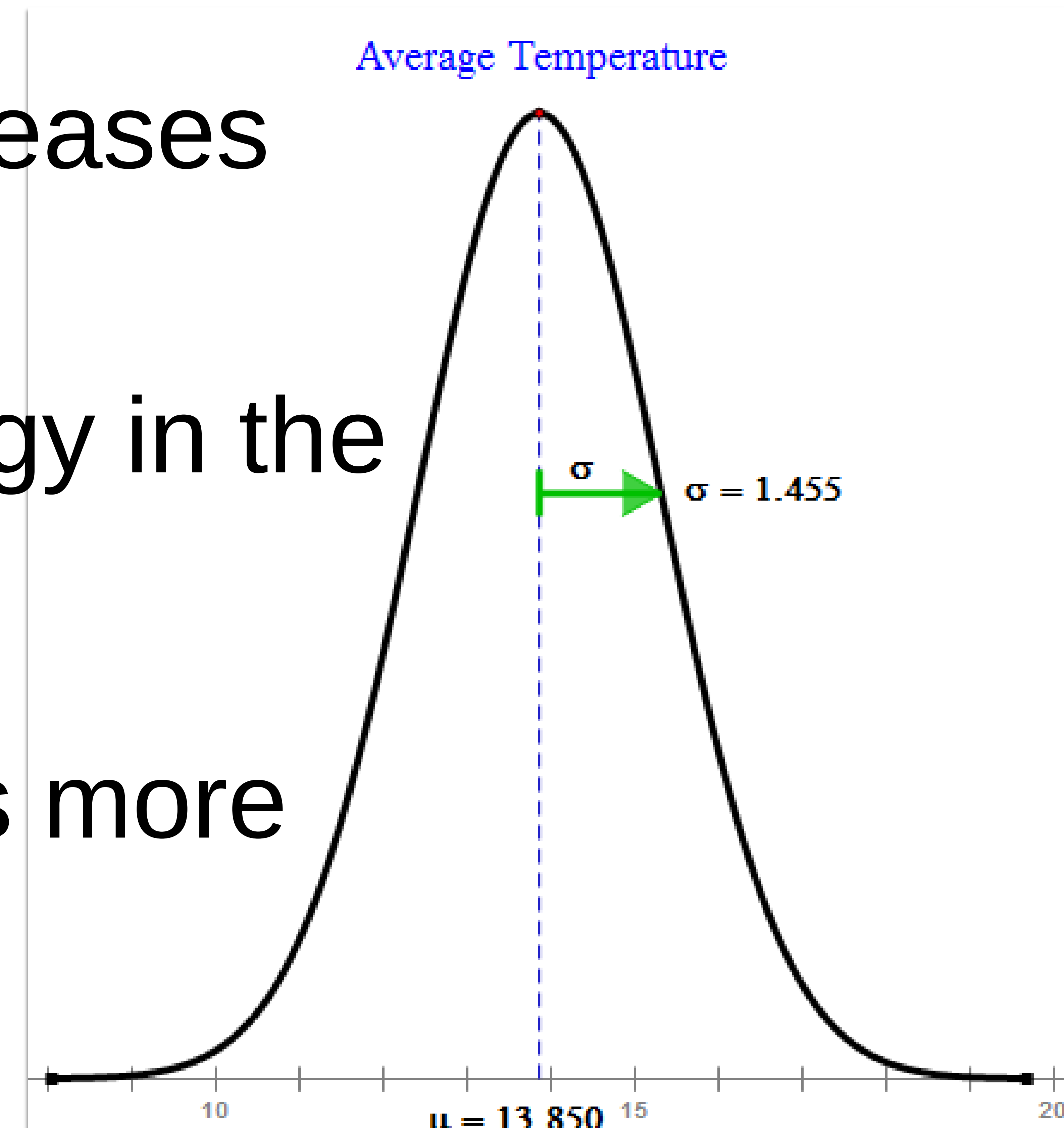
Applications – Explaining Climate Change

What happens when average temperature increases?

Probability of higher temperatures increases

Higher temperatures mean more energy in the atmosphere

More energy in the atmosphere means more violent storms



Probability

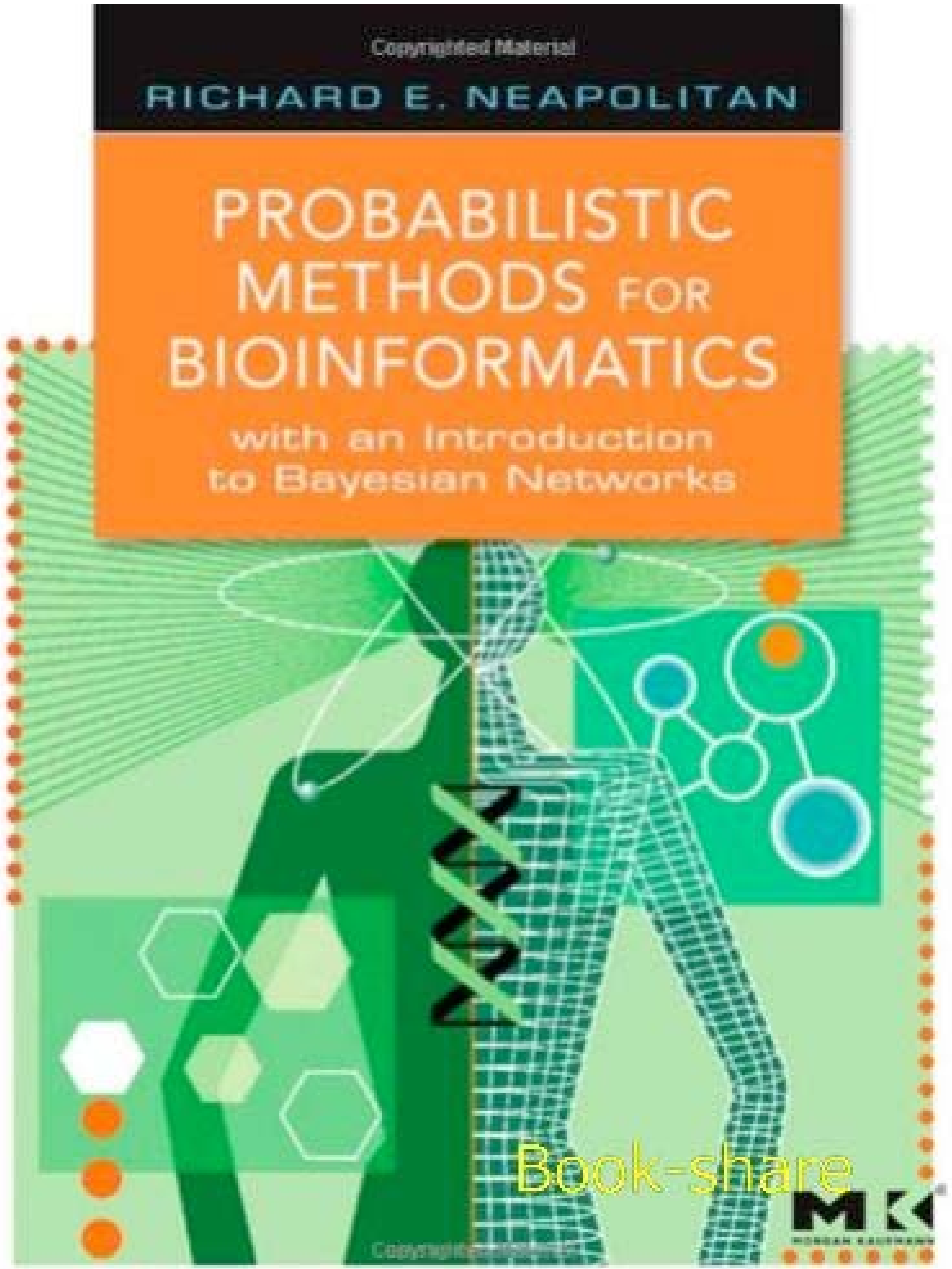
Applications - Finance

Two thirds of all money traded on the stock exchange is now done by computers using statistical models.

Access to these models is difficult due to the highly secret nature of the research - there is alot of money to be made!

Probability

Applications - Medical Research



Probabilities enter almost every aspect of medical treatment and medical research.

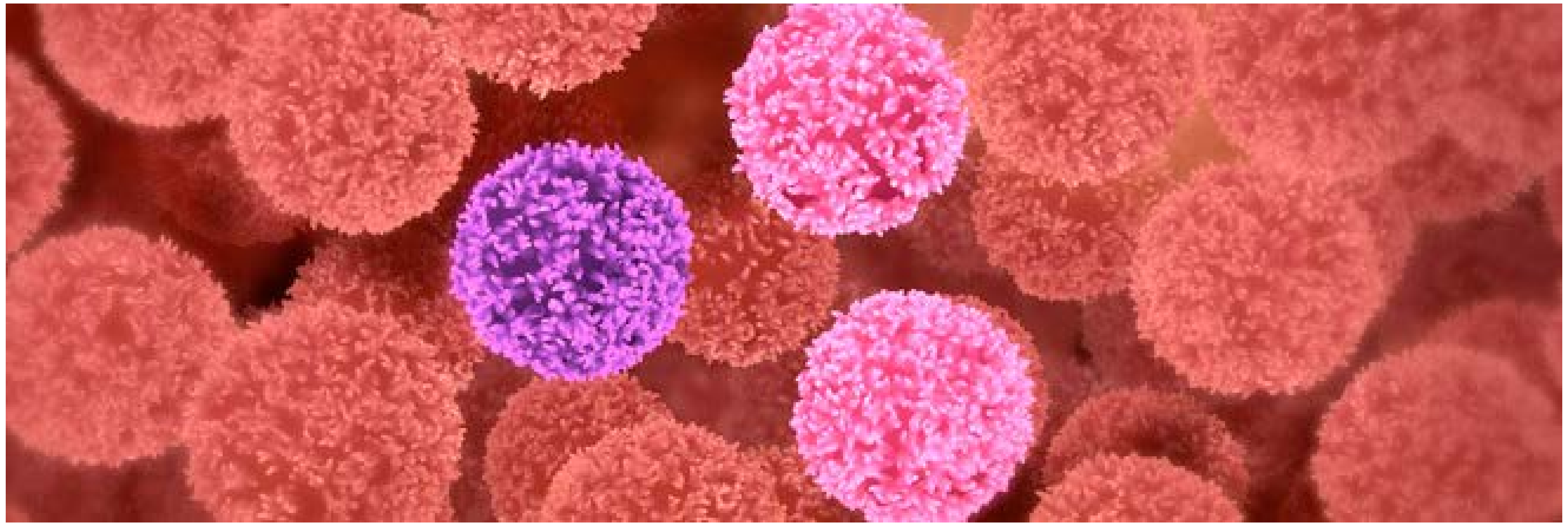
http://www.downeu.net/s/Social+Networks+and+Health++Models+Methods+and+Applications

Probability

Bioinformatics

For students with maths and programming skills there is a new career that might interest them...

Bioinformatics mixes mathematics, biology and programming.



The demand is very high.

http://www.wehi.edu.au/faculty/disease_research/

Probability

Bioinformatics

Through the UROP program, undergraduate students are employed in biomedical research, preparing 'work ready' graduates in this emerging field.

They work with real scientists and mathematicians throughout their course, gaining useful employment, experience and a taste of what the work is like.

http://www.bio21.com.au/pages/urop

Probability

Bioinformatics

The Walter and Eliza Hall Institute (WEHI) at the University of Melbourne and the Royal Melbourne Hospital has a Bioinformatics Division that investigates:

Cancer - Breast Cancer, Leukaemia and Lymphoma, Multiple Myeloma

Chronic inflammatory diseases - Coeliac Disease, Type 1 Diabetes, Rheumatoid Arthritis, Transplantation

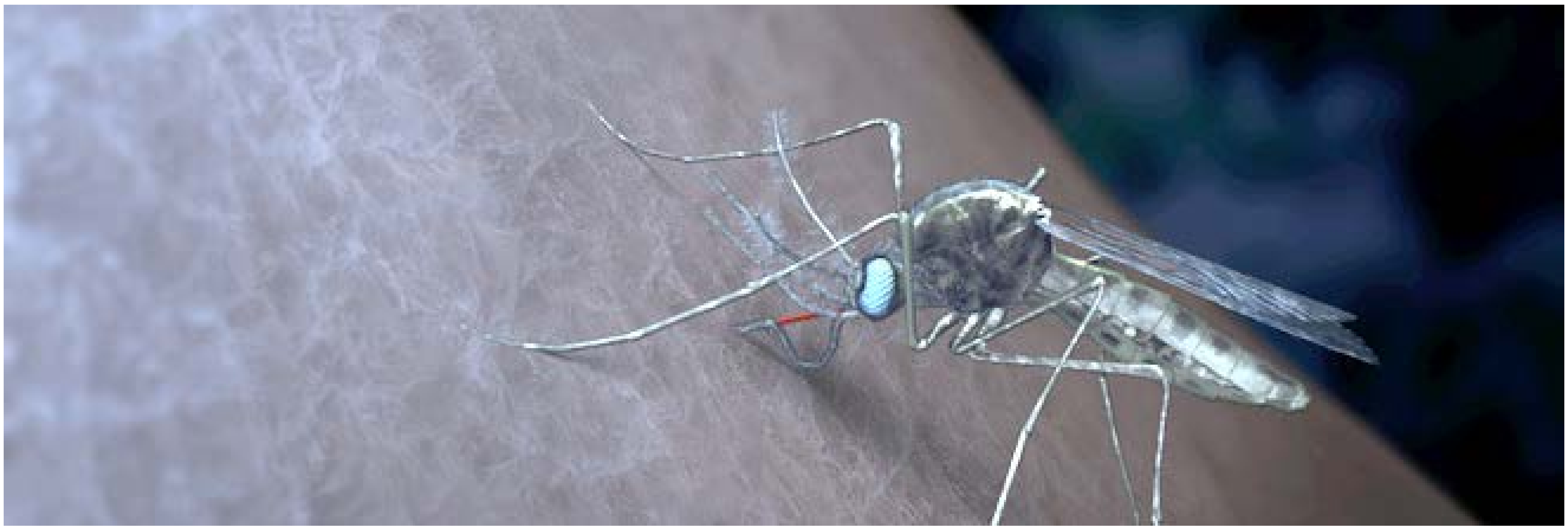
Infectious diseases - Malaria, Chronic Infectious Diseases

http://www.wehi.edu.au/faculty/disease_research/

Probability

Bioinformatics and Malaria

A child dies of malaria every 15 seconds and the total number of deaths every year is 1.5 to 2.7 million.

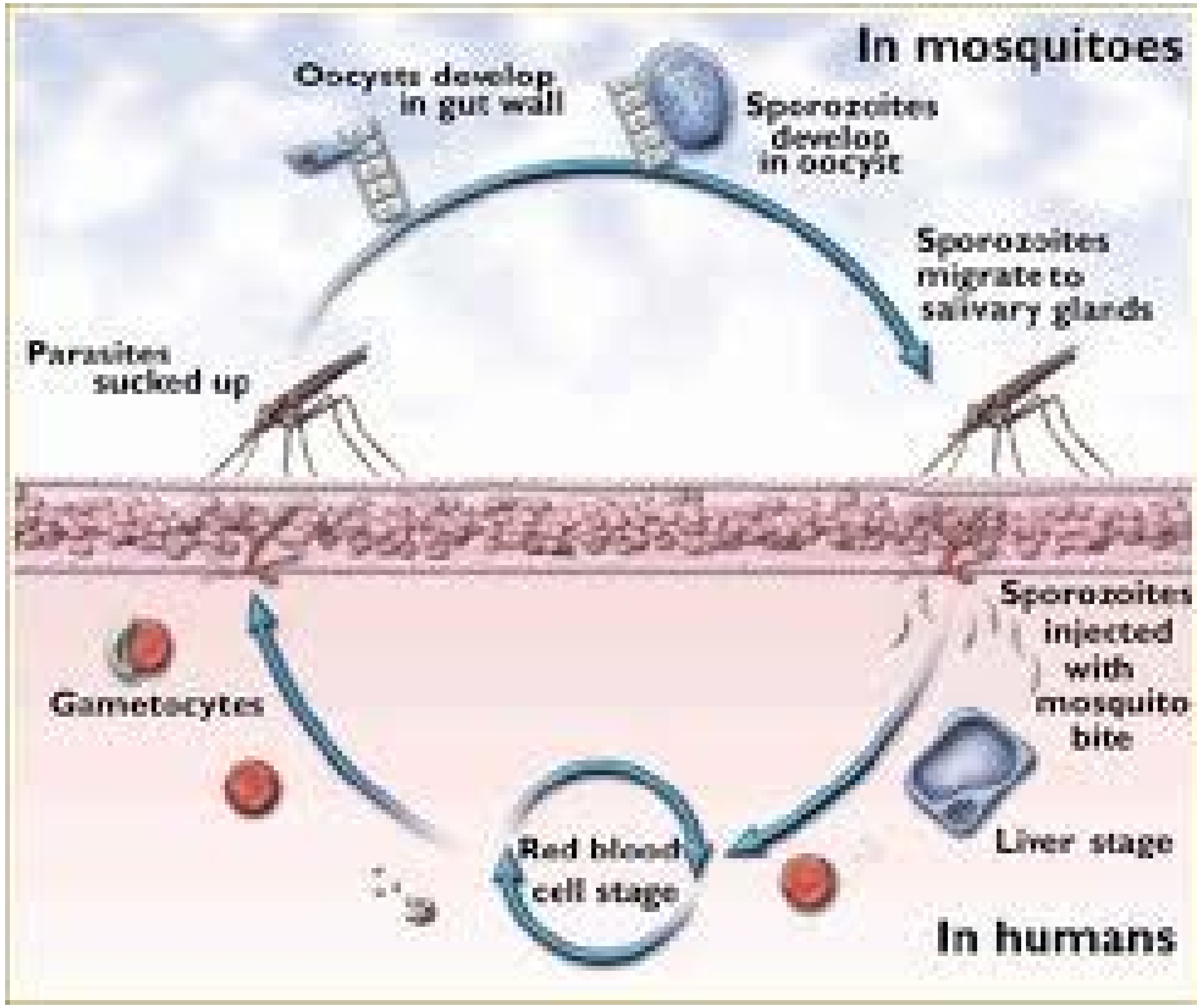


http://www.wehi.edu.au/faculty/disease_research/

Probability

Bioinformatics and Malaria

By comparing the protein sequences in two similar parasites we can know more about how each of them works.



http://www.malariaworld.org/category/continent/africa?page=4

Probability

Bioinformatics and Malaria

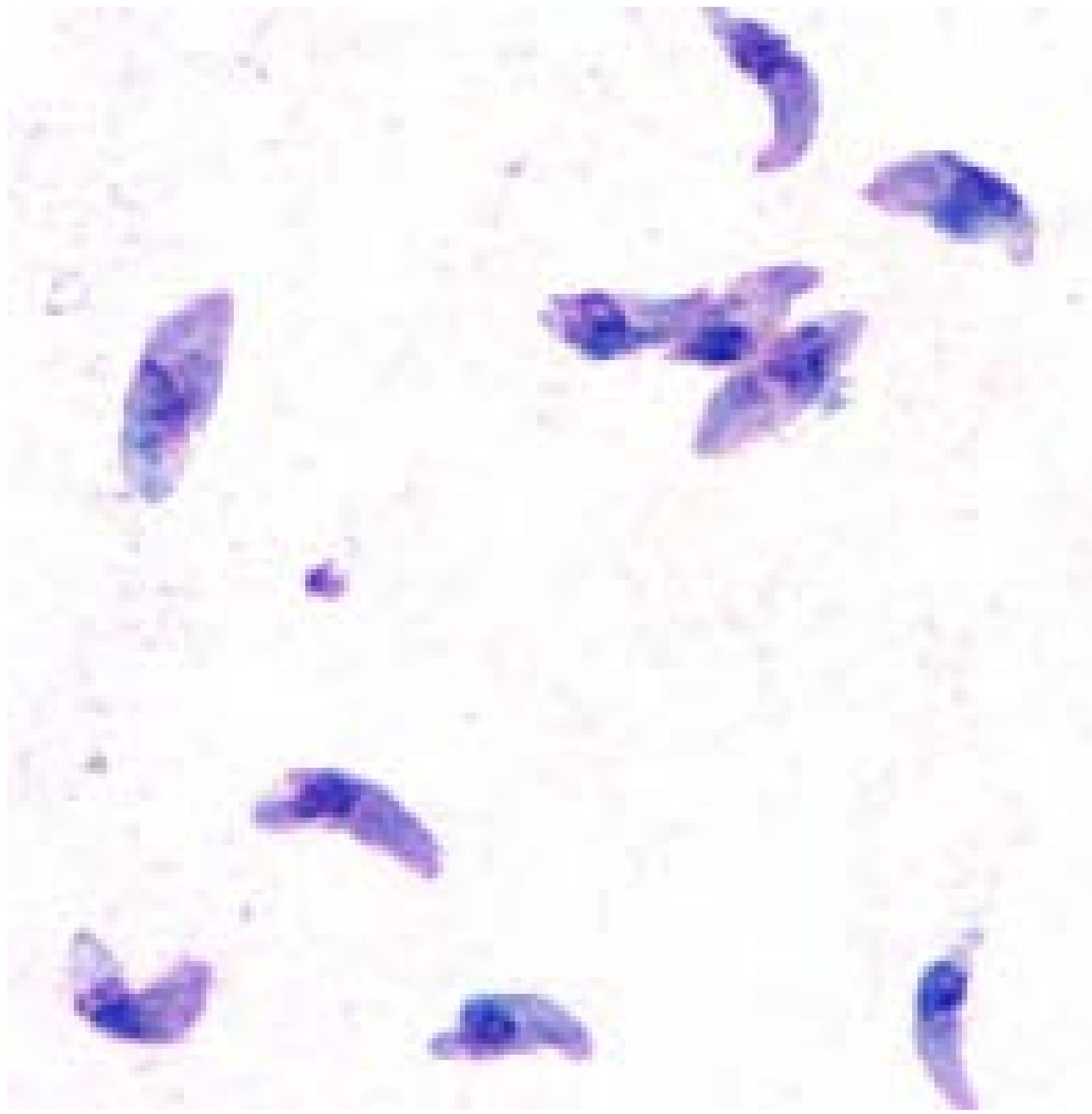
Plasmodium falciparum is a protozoan parasite, one of the species of Plasmodium that cause **malaria** in humans.

http://en.wikipedia.org/wiki/File:Plasmodium_falciparum_01.png

Probability

Bioinformatics and Malaria

Toxoplasma gondii is a species of parasitic protozoa.



Usually transmitted by cats, **Toxoplasmosis** is usually minor and self-limiting but can have serious or even fatal effects on a fetus whose mother first contracts the disease during pregnancy or on an immunocompromised human or cat.

http://en.wikipedia.org/wiki/Toxoplasma_gondii

Probability

Bioinformatics and Malaria

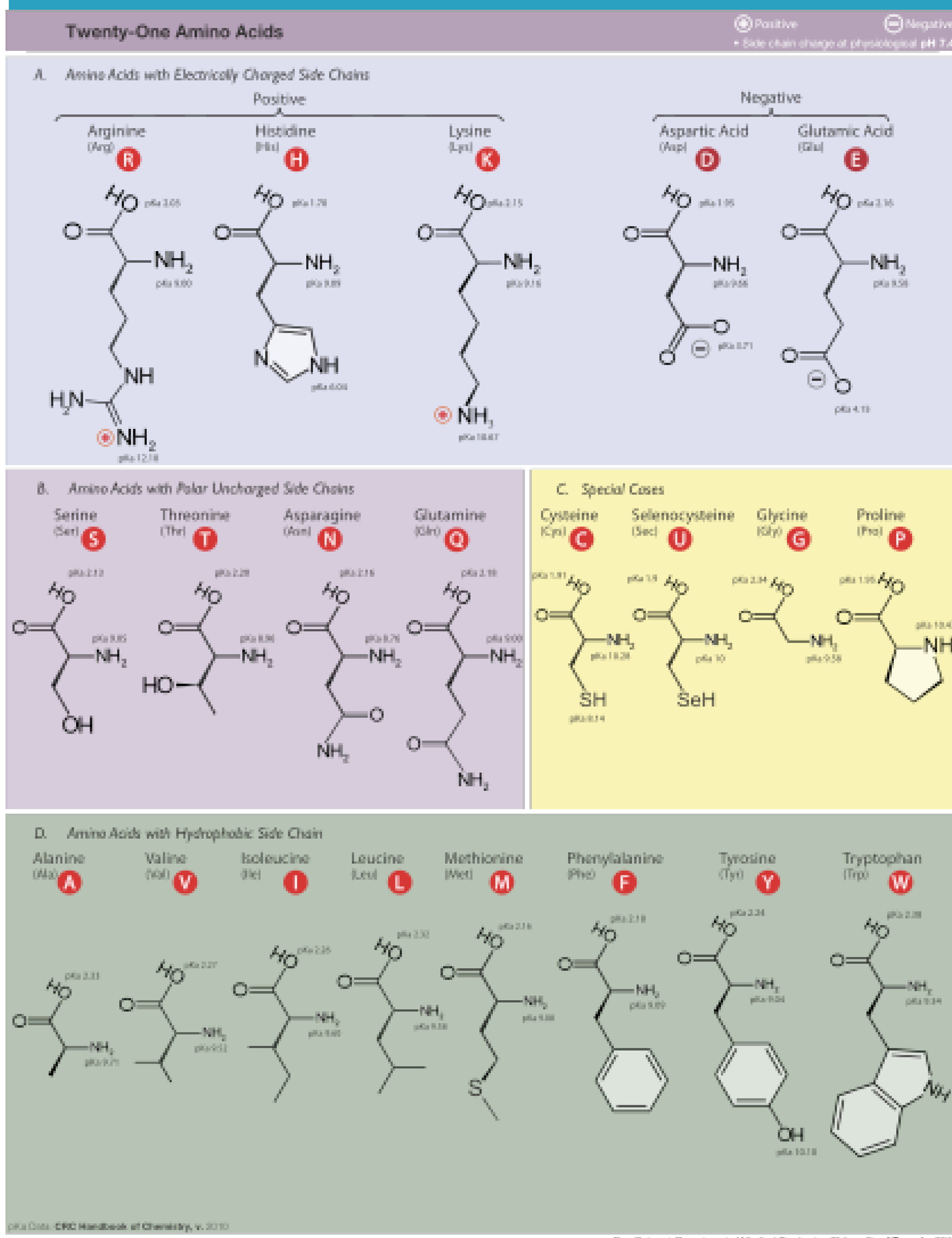
Hopefully some of what we know about toxoplasma can help us understand more about malaria.

Everything in a cell is made of proteins. Proteins are used for transport within cells and are part of the cell structure.

Knowing about the similar pathway proteins shared by these two species, we can use similar methods to transport drugs into a malaria cell.

Probability

Bioinformatics and Malaria



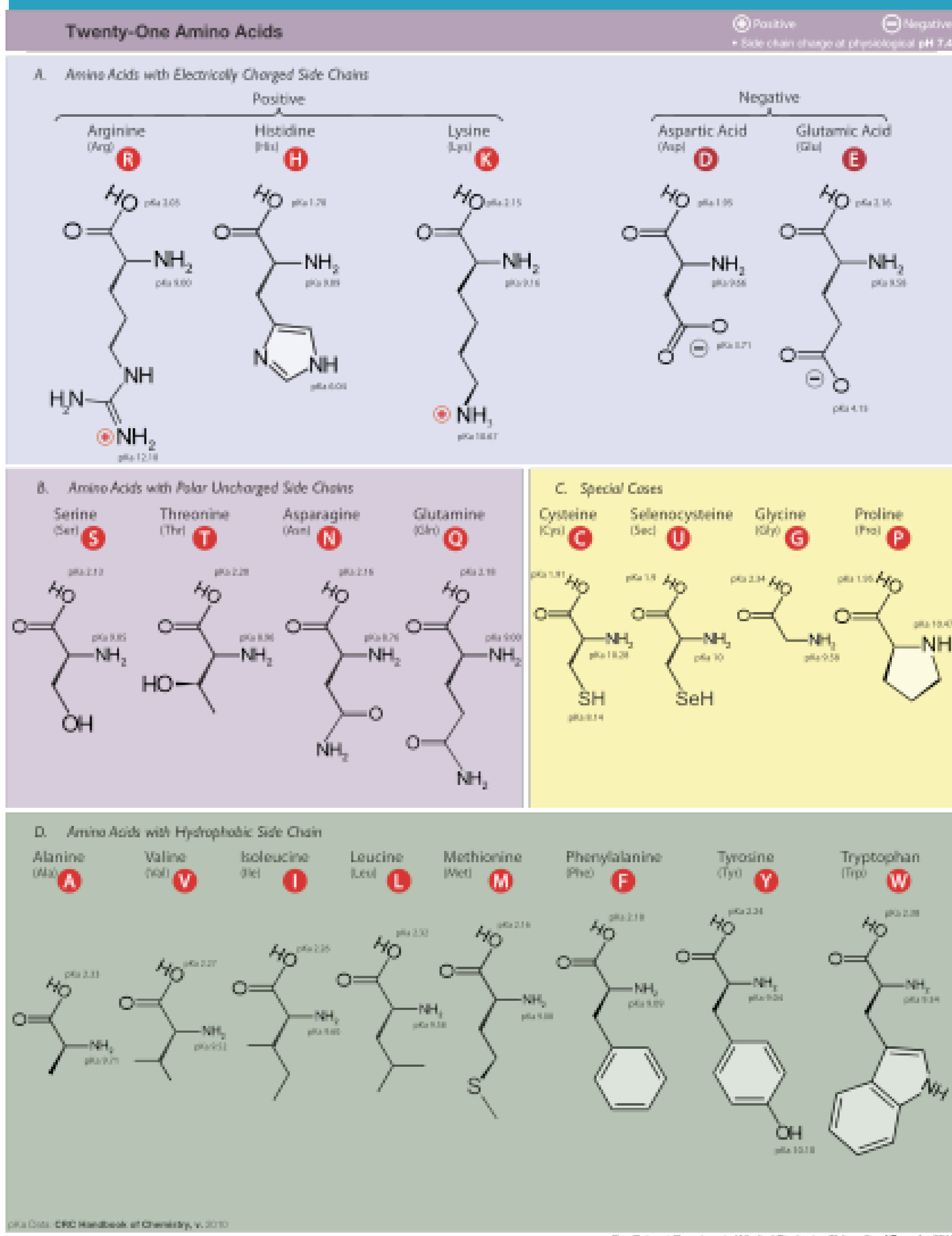
There are 21 Proteins that make up all genetic material.

Denoted by the letters:

RHKDESTNQCUGPAVILMFYW

Probability

Bioinformatics and Malaria



Predicting the likelihood of finding the same arrangement of proteins... using Python Programming software.

make_alignments.py

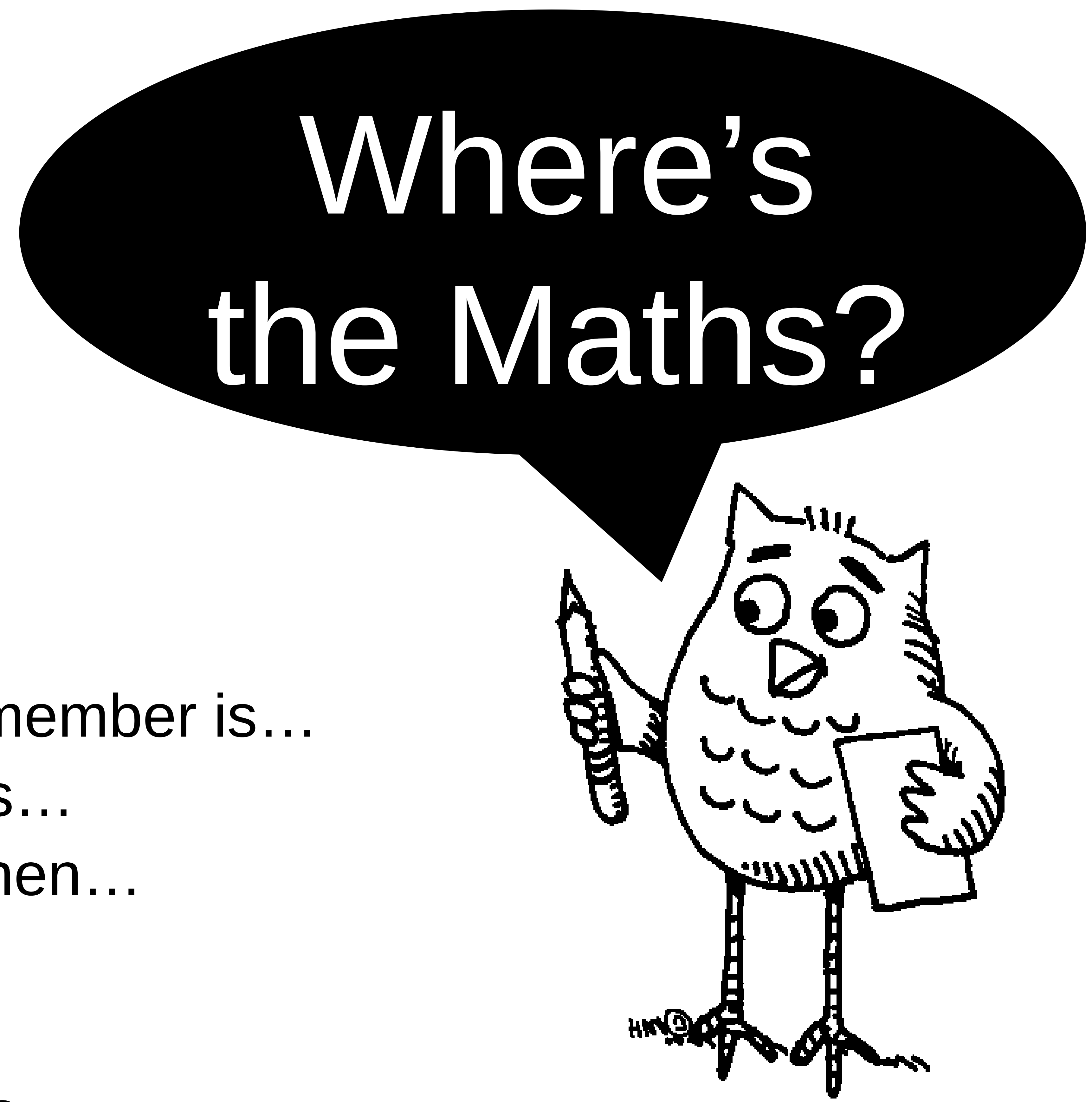
Last Saved: 11/29/11 3:44:24 PM
File Path: ~/Documents/proj1b/make_alignments.py

make_alignments.py FastNW

```
    return scormat, arrow

def Backtrace( arrow, seq1, seq2):
    st1, st2 = ","
    v,h = arrow.shape
    ok = 1
    v -=1
    h-=1
    while ok:
        if arrow[v,h] == 0:
            st1 += seq[v-1]
            st2 += '-'
            v -=1
        elif arrow[v,h] == 1:
            st1 += '-'
            st2 += seq2[h-1]
            h-=1
        elif arrow[v,h] == 2:
            st1 += seq1[v-1]
            st2 += seq2[h-1]
            v-=1
            h-=1
        if v == 0 and h==0:
            ok = 0
    #reverse the strings
    st1 = st1[::-1]
    st2 = st2[::-1]
    return st1,st2

def FastNW(subvalues


def align(seqs,protein_seq):
    from numpy import *

    return None

def print_nicely():
    data = open('/Users/lmcintosh/Documents/proj1b/DATA.txt','r')
    output = open('/Users/lmcintosh/Documents/proj1b/output.txt','w')
    count = 0
    for row in data:
        i = row.split('\t')
        if i[6] == 'y':
            count +=1
        if (count == 200 and i[6]=='y'):
            protein_seq = open('/Users/lmcintosh/Documents/proj1b/clustal_data/%s.aln' %i[0],'r')
            length = int(find_start(i[0]))
            seqs = i[9].strip('"').split(',')
            j = 0
            line1 = []
            line2 = []
            line3 = []

            for line in protein_seq:
                j+=1
                #first output is on fourth line
                #=> everything on 4+3+n4 is useless
                if j > 3:
                    if (j-4)%4 == 0:
```

What is a mathematics curriculum?

We already knew…
We remembered…
We used equipment…
We need to find out…
It was interesting when…
The tricky bit was…
We didn't know that…
It was cool when…
The important thing to remember is…
A new word we learnt was…
Our group worked well when…
We discovered…
Congratulations to…
The strategy we used was…



We thought…
We said…
We wrote…
We saw…
We heard…
We know…
We drew…
We said…
We asked…
We felt…
We liked…
We learnt…
We didn't like…
We found out…