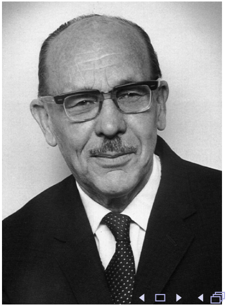
Topic4: Probability, Random Variables and Distributions

Outline

Example: Coin Tossing during WWII

John Edmund Kerrick (1903–1985) was a mathematician noted for a series of experiments in probability which he conducted while interned in Nazi-occupied Denmark (Viborg, Midtjylland) in the 1940s.

Kerrich had travelled from South Africa to visit his in-laws in Copenhagen, and arrived just 2 days after Denmark was invaded by Nazi Germany!



Fortunately Kerrich was imprisoned in a camp in Jutland run by the Danish Government in a ‘truly admirable way’.

With a fellow internee Eric Christensen, Kerrich set up a sequence of experiments demonstrating the empirical validity of a number of fundamental laws of probability.

- They tossed a coin 10,000 times and counted the number of heads.
- They made 5000 draws from a container with 4 ping pong balls (2x2 different brands), 'at the rate of 400 an hour, with - need it be stated - periods of rest between successive hours.'
- They investigated tosses of a 'biased coin', made from a wooden disk partly coated in lead.

In 1946 Kerrich published his finding in a monograph, *An Experimental Introduction to the Theory of Probability*.

Kerrich1950Paper

How many heads do you think he counted? What is the probability of getting a head on a fair coin?

Why study Probability?

Probability is often perceived as ‘scary’ because it conjures up impossible problems from School Maths Competitions! However ...

1. Probability is an essential part of every day life.
   “The most important questions of life ... are indeed for the most part only problems of probability.”
   (Laplace, *Théorie Analytique des Probabilitiés*, 1814) ▸ Laplace1
2. Probability Theory makes many school problems easy.
   “The theory of probabilities is basically just common sense reduced to calculus.”
   (Laplace, *Essai philosophique sur les probabilités*, 1814) ▸ Laplace2
3. Probability (or the p-value) is crucial for decision making in Hypothesis Testing, which is essential to scientific research (Part 3).

What is Probability?

Definition (Probability)

The probability of an event is a measure of the likelihood of that event occurring.

Probability Theory is a set of mathematical tools which dates back centuries to casino type games. A more formal theory was developed in the 1930s by the Russian mathematician A. N. Kolmogorov. ▸ History Video ▸ Appendix

Three types of Probability

(1) Subjective probability: based on belief
The probability of an event is based on the strength of one's belief.

Tossing a Fair Coin

If I toss a fair coin once, the probability of getting a head is . . .

We rely on subjective probability for everyday decisions. But it can be abused.

On March 27 1977 a PanAm 747 jet and a KLM 747 jet collided on an airport runway in the Canary Islands killing 583 people. Both jets had been scheduled for the Las Palmas Airport, but were diverted to Los Rodeos Airport (now called Tenerife) after a group of militants set off a small bomb at the airport's flower shop earlier that day.

The following statement was released:

NEW YORK, Mon: Mr. Webster Todd, Chairman of the American National Transportation Safety Board said today statistics showed that the chances of two jumbo jets colliding on the ground were about 6 million to one. (AAP, quoted in West Australian, 1977)

Australian statistician Terry Speed questioned the Chairman, with the following response:

Dear Professor Speed,
In response to your aerogram of April 5, 1977, the Chairman's statement concerning the chances of two jumbo jets colliding (6 million to one) has no statistical validity nor was it intended to be a rigorous or precise probability statement. The statement was made to emphasize the intuitive feeling that such an occurrence indeed has a very remote but not impossible chance of happening. Thank you for your interest in this regard.

(2) Frequentist (or simulation) probability: based on data
The probability of an event is the proportion of times that event would occur in a large number of repeated experiments (simulation).

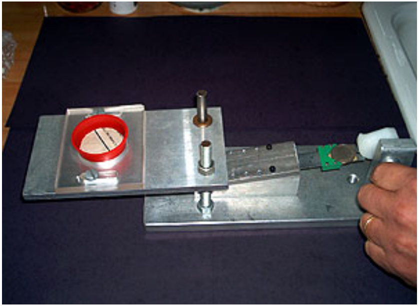
In some ways, subjective probability is an informal version of frequentist probability, using one's own history and experience ('personal data') as the basis.

Tossing a Fair Coin

The probability of tossing a fair head $= P(H) =$ The long run proportion of heads if we toss a fair coin a large number of times.

Famous Coin Tossers

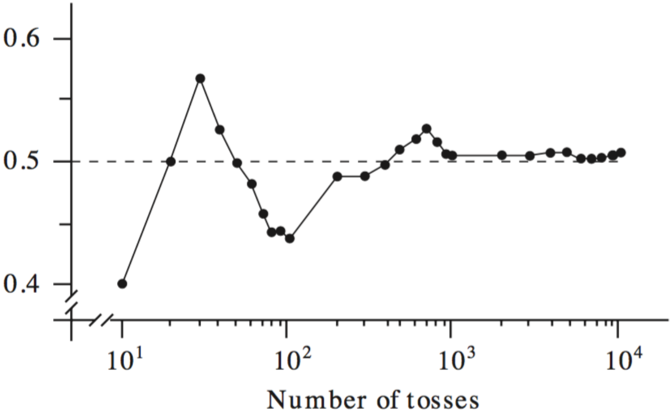
Person	Number of Tosses n	Number of Heads x	$P(Head)$
Count Buffon (1707-1788) ▸ Buffon	4040	2048	0.507
Karl Pearson (1857-1936) ▸ Pearson	24000	12012	0.5005
John Kerrick (1903-1985 war camp) ▸ Kerrick	10000	5067	0.5067
Perci Diaconis (1945-present) ▸ Diaconis	machine		0.51



In R Simulations

```
table(sample(c("H","T"),4040,replace=T))/4040

##
##         H         T
## 0.4933168 0.5066832

table(sample(c("H","T"),24000,T))/24000

##
##         H         T
## 0.5015833 0.4984167

table(sample(c("H","T"),10000,T))/10000

##
##     H     T
## 0.503 0.497
```

Kerrich's Tosses: Proportion of Heads



This illustrates the 'Law of Averages' or 'Law of Large Numbers': the proportion of heads becomes more stable as the length of the simulation increases and approaches a fixed number called the 'relative frequency'.

(3) Theoretical (or classical) probability: based on model
The probability of an event is based on a model of the context.

Two models for Coin Tossing:

(1) Physics model: the side that the coin lands on is determined by a number of complicated factors such as which way up it started, the degree of spin, the speed and angle with which it left the thumb and how far it has to fall. PercisDiaconisVideo1 PercisDiaconisVideo2

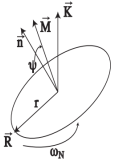


Figure 2: Coordinates of Precessing Coin.

(2) Standard model: Tossing a fair coin results in a head and a tail half the time, in an unpredictable order (random).

Tossing a Fair Coin

Given a coin toss has 2 outcomes, the probability of tossing a fair head $= \frac{1}{2}$.

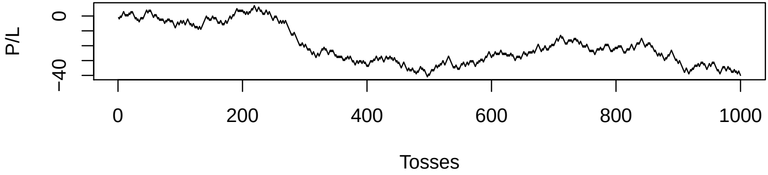
This is the model used for decisions in sport (eg which team starts with the ball in AFL, or which team bats or bowls first in cricket).

Extension

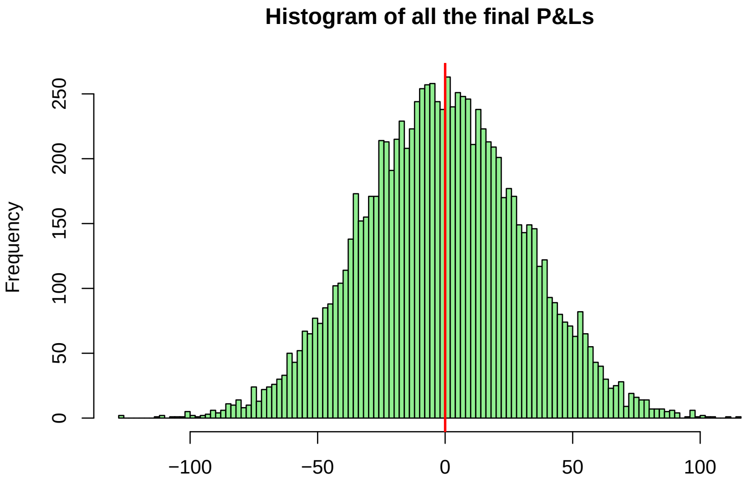
Suppose you toss a fair coin 1000 times: when it lands heads you receive $1, and when it lands tails you pay me $1. What is your expected profit or loss?

Using Simulation:

```
set.seed(1)
CoinTosses <- sample(c(-1,1), 1000, replace = TRUE)
plot(cumsum(CoinTosses),xlab="Tosses", ylab="P/L", type="l'
```



We could run this simulation 10000 times, resulting in:

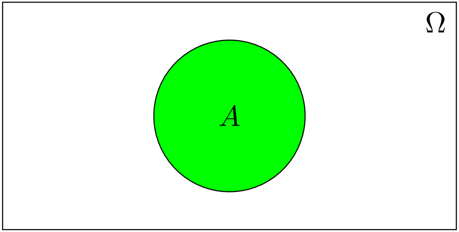


In Topic 5, we will see how to solve this theoretically.

Describing Simple Probability Models

We can describe simple probability models using set notation and the Venn diagram.

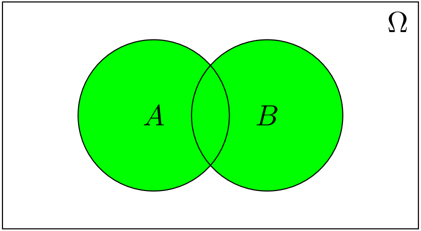
Symbol	Name	Meaning
Ω	Sample Space	Everything that can occur.
$A, B, \ldots$	Events	A subset of the sample space.
$\emptyset$	Empty Set	An event which cannot occur.
A' or A^c	Complement	Everything not in A.
$A \subset \Omega$	Belongs to	Event A belongs to sample space Ω.



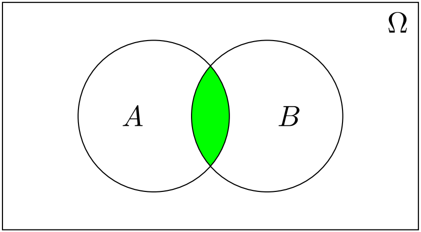
$A \subset \Omega$

If A and $B \subset \Omega$:

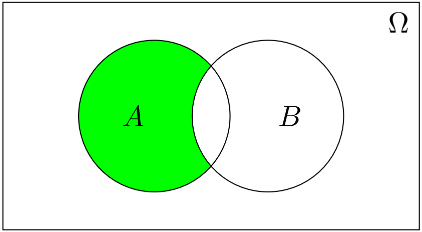
Symbol	Name	Meaning
$A \cup B$	Union	A or B or both occur.
$A \cap B$	Intersection	Both A and B occur.
$A \setminus B$	Minus or Relative Complement	In A but not in B.



$A \cup B$



$A \cap B$



$A \setminus B$

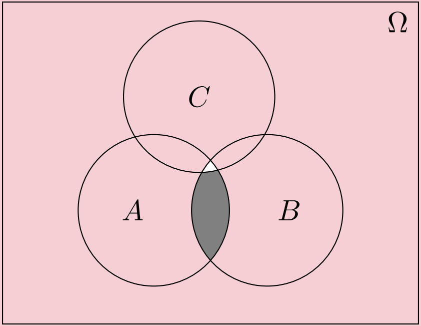
If $A_1, A_2, \ldots \subset \Omega$:

Symbol	Name	Meaning
$\cup A_{i=1}^n$	Union	One or more of A_i occur.
$\cap A_{i=1}^n$	Intersection	All A_i occur.

Example

Your Turn (Set Notation)

Describe the grey and white regions in set notation.



Your Turn (Sample Space)

- A coin is tossed to start a cricket match. What is the probability of landing on its edge?
- For the following scenarios, what is Ω?
  - A coin is tossed until a head occurs.
  - The 2016 Australian census question about Gender.
  - Measuring tomorrow's rainfall.

Probability Results

- Classical Probability
  Where the sample space Ω consist of a finite, known number of equally likely outcomes (eg coins, dice, cards), the probability of an event $A \subset \Omega$ occurring is

$$\boxed{P(A) = \frac{\text{Number of ways } A \text{ can occur}}{\text{Total number of possible outcomes in } \Omega}}$$

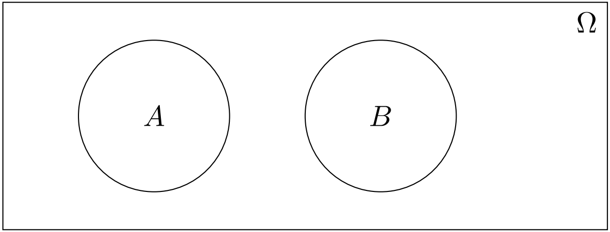
  where
  $P(A) = 0$ (impossible event)
  $P(A) = 1$ (certain event)

- Complement
  The probability that an event A does not occur is

$$\boxed{P(A') = 1 - P(A)}$$

- Mutually Exclusive
  2 events A and B are mutually exclusive if

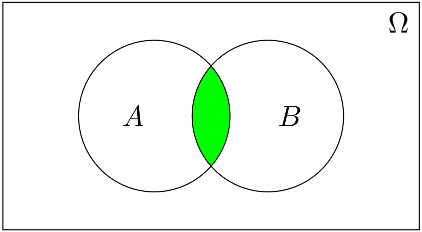
$$\boxed{P(A \cap B) = 0}$$



Note: This is not a 'rule' but a definition. We position it here, to see comparison with independence

- Independence
  2 events A and B are independent (the result of one does not affect the result of the other) iff

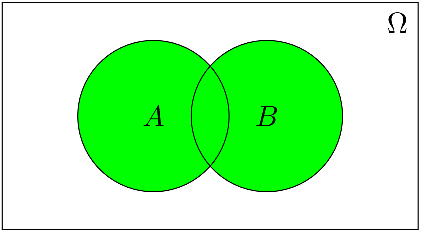
$$\boxed{P(A \cap B) = P(A)P(B)}$$



$P(A) = 0.2$, $P(B) = 0.5$, $P(A \cap B) = 0.1$

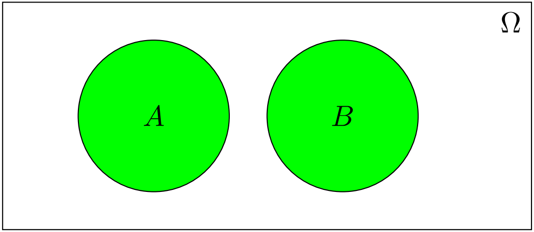
- Union Rule
  For events A and $B \subset \Omega$, then

$$\boxed{P(A \cup B) = P(A) + P(B) - P(A \cap B)}$$



For mutually exclusive events, this reduces to

$$\boxed{P(A \cup B) = P(A) + P(B)}$$



- Conditional Probability
  For events A and $B \subset \Omega$ where $P(B) \neq 0$, the conditional probability of A given B is

$$\boxed{P(A|B) = \frac{P(A \cap B)}{P(B)}}$$

This leads to a second definition of independence:
2 events A and B are independent iff

$$\boxed{P(A|B) = P(A)}$$

- Bayes Theorem

$$\boxed{P(A|B) = \frac{P(B|A)P(A)}{P(B)}}$$

Bayes Theorem

- The Law of Total Probability

$$\boxed{P(A) = \sum_i P(A|B_i)P(B_i)}$$

for any i for which $P(B_i) \neq 0$.

Counting Short Cuts

Computing classical probabilities amounts to careful counting, so we have the following short-cuts.

Multiplication Principle for k stage problems

If an experiment consists of k stages where the ith stage has n_i outcomes (for $i = 1, 2, \ldots k$), then the total number of outcomes is $n_1 n_2 \ldots n_k$.

Definition (Permutations, Factorials)

Any x distinct objects can be permuted (or rearranged) in $x! = x(x-1)(x-2)\ldots 321$ ways.

Definition (Combinations, Binomial Coefficients)

The number of ways of choosing a group of x objects from n objects is

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

where by definition $0! = 1$.

```
factorial(10)/(factorial(8)*factorial(2))

## [1] 45

choose(10,2)

## [1] 45
```

Examples

Your Turn

- Suppose a mobile phone number of length 10 starts with the numbers 04. How many unique numbers can be issued? Assuming all numbers are randomly issued, what is the probability of getting the number $04********$, where $* =$ 0,1, ... 9.
- A chain is formed from n independent links, with a probability of θ that any link fails under a specified load. What is the probability that the chain fails under the load?
- The probability that a component lasts at least x hours is $e^{-x/100}$, for $x > 0$. What is the probability that the component lasts at least 10 hours, given it has already lasted at least 6 hours?

Why Random Variables?

Definition (Random Variable)

A random variable is a real valued function on a sample space Ω, that captures both the outcomes and associated probabilities.

A random variable is a simple way of summarising a probability problem. It sharpens our focus by summarising the outcomes and highlighting the events of interest.

We use a capital Roman letter X to denote a random variable and the corresponding lower case letter x to denote the values of the random variable. We will consider discrete and continuous random variables, depending on the underlying context.

Examples

Your Turn

Consider the following 3 functions. Is X a random variable?

x	-1	0	1	Total
$f(x)$	-0.2	1	0.2	

x	-1	0	1	Total
$f(x)$	0.4	0.2	0.3	

x	-1	0	1	Total
$f(x)$	0.4	0.3	0.3	

Coin Tossing 5 times

Toss a fair coin 5 times. What is the probability of getting 4 or more heads?

Long way: List Sample Space
$\Omega = \{HHHHH, \ldots, TTTTT\}, |\Omega| = 2^5 = 32$
$P(\text{4 or more heads}) = \frac{6}{32} = \frac{3}{16}$

Neater way: Summarise by defining a random variable
Let $X =$ the number of heads in 5 tosses of a coin, $x = 0, 1, \ldots, 5$.

x	0	1	2	3	4	5
$P(X = x)$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

$$P(X \geq 4) = \frac{6}{32}$$

Your Turn (Die)

Toss a fair die 2 times. What is the probability of getting a total of 8 or more?

Long way: List Sample Space

$$\Omega = \left\{ \begin{array}{c} \text{11,12,13,14,15,16} \\ \text{21,22,23.24.25.26} \\ \text{31,32,33,34,35,36} \\ \text{41,42,43,44,45,46} \\ \text{51,52,53,54,55,56} \\ \text{61,62,63,64,65,66} \end{array} \right\}, |\Omega| = 36$$

$$P(\text{Total of 8 or more}) = \frac{15}{36}$$

Neater way: Summarise by defining a random variable

Let X = the total of 2 tosses of a die, $x = 2, 3, \ldots, 12$.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

$$P(X \geq 8) = \frac{15}{36}$$

What is a Distribution?

Definition (Distribution)

A distribution defines the behaviour of a situation modelled by a variable X, in particular

- the set of all possible values $\{x\}$
  "what can happen"
- the probability of each value (discrete) or each range of values (continuous) occuring.
  "how often everything happens"

Probability Functions

Definition (Probability Functions)

For a variable X, we can describe the probabilities by:

- the probability distribution function (for discrete variables)
  $P(X = x)$
- the probability density function (pdf) (for continuous variables)
  $f(x)$
- the cumulative distribution function (CDF) (for both discrete and continuous variables)
  $F(x) = P(X \leq x)$

Types of Distributions

Definition (Types of Distributions)

We can characterise distributions in 2 ways:

By context:

- population distribution
- sample distribution
- sampling distribution (for a statistic like $\bar{X}$).

By underlying nature:

- discrete distribution
- continuous distributions

Notes on Distributions

- The distribution can be referred to as the *probability* distribution or the *statistical* distribution.
- The sample distribution can suggest the population distribution, especially for large sample size n.
- For a population distribution, the mean is μ and the variance is σ^2. For a sample distribution, the mean is $\bar{x}$ and the variance is s^2.

	Population: Parameter	Sample: Statistic
Mean	μ	$\bar{x}$
Standard Deviation	σ	s