Probability theory and statistics

Dr Gianluca Campanella

10th May 2016

Contents

Probability theory

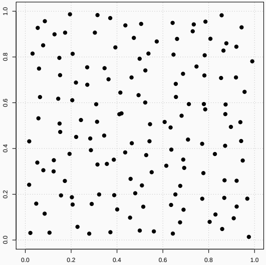
Random variables

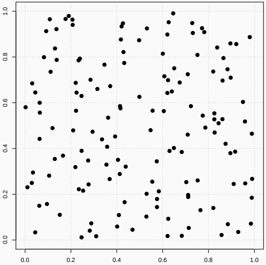
Probability distributions

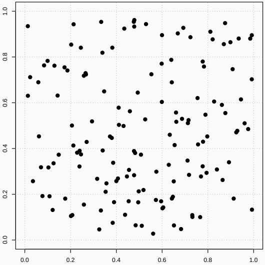
Statistics

Probability theory

Exercise: ‘random’ points







What is probability?

> *'The extent to which something is likely to happen or be the case'*
>
> — *Oxford English Dictionary*

Examples

- Probability that it will rain tomorrow
- Probability that you will win the lottery

Example: sources of uncertainty

Imperfect information

Current predictive tools can only assigns a number between 0 and 1 indicating our degree of certainty

Stochastic process

The experiment is designed to produce uncertain results

Probability theory

What?

The branch of mathematics concerned with the analysis of **random phenomena**

How?

Using mathematical **abstractions** of measured quantities and non-deterministic events

Why?

To identify **patterns** in (apparently) random occurrences

Statistical regularity

We cannot predict with certainty
if it's going to rain tomorrow

but

we can predict 'averages'

(Average rainfall in London in May: 44.9 mm)

Statistical regularity

In summary...

- Probability theory predicts the behaviour of random phenomena **in the long run**
- If this information is useful, probability can be a valuable tool for **decision-making**

Random variables

Random variables

- 'Encapsulate' random events
- Mathematically, they are functions mapping the **sample space** to some numerical property (discrete or continuous)

Notation

- X, Y, ... (upper case) are random variables
- $X = x$ (lower case) is a value (**realisation**) of X
- $\Pr(X = x)$ is the probability that $X = x$

Random variables

- ‘Encapsulate’ random events
- Mathematically, they are functions mapping the **sample space** to some numerical property (discrete or continuous)

Example

- X represents the event ‘the UK leaves the EU’
- $X = 1$ represents the UK leaving the EU
- $\Pr(X = 1)$ is the probability that the UK leaves the EU

Types of random variables

Discrete

- Countable (usually finite) outcomes
- Each has a (non-zero) probability assigned

Continuous

- Infinite outcomes
- Not possible to assign a non-zero probability to each

Exercise: discrete random variable

A fair die

x	$\Pr(X = x)$
1	1/6
2	1/6
3	1/6
4	1/6
5	1/6
6	1/6

Maximum of two fair dice

- How many outcomes?
- $\Pr(X = 1)$?
- $\Pr(X = 6)$?

Probability distributions

Probability distributions

Simplified approximations to reality

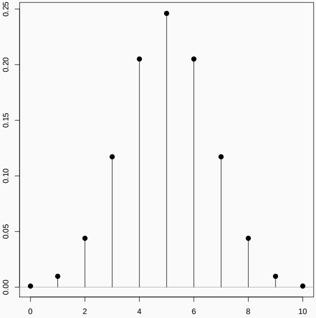
- Detailed enough to capture important characteristics and serve as **prediction tools**
- Simple enough to be usable in practice

Specifying probability distributions

Discrete RV

$\downarrow$

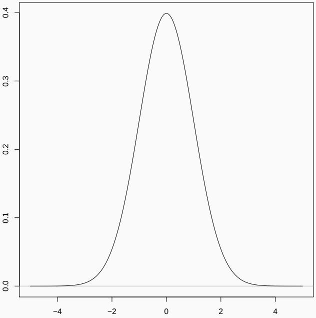
Probability mass function



Continuous RV

$\downarrow$

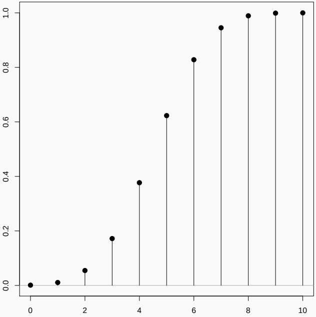
Probability density function



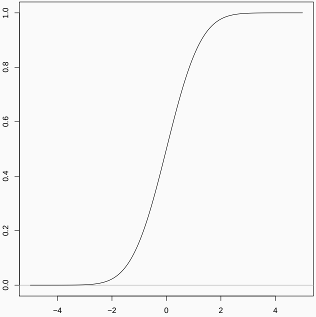
Specifying probability distributions

Cumulative distribution function

Discrete RV



Continuous RV



Something always happens...

$$\sum f_X(x) = 1 \quad \text{or} \quad \int f_X(x)\ \mathrm{d}x = 1 \quad \text{or} \quad \lim_{x \to \infty} F_X(x) = 1$$

Notation

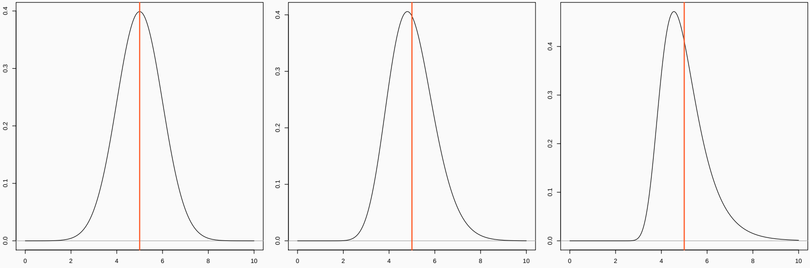
- f_X is a probability function
- $F_X = \Pr(X \leq x)$ is a cumulative distribution function

Characterising probability distributions

- Central tendency
    - Mean (expected value)
    - Median
    - Mode
- Dispersion
    - Variance
    - Minimum and maximum
    - Quantiles

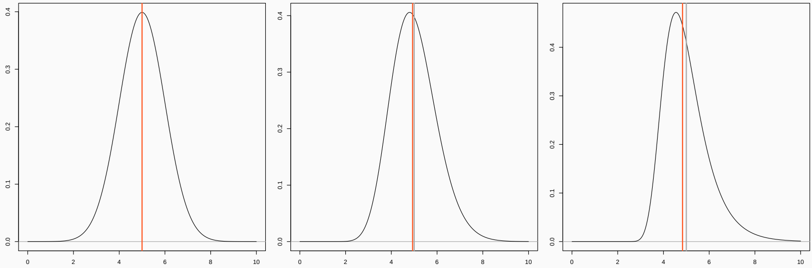
Measures of central tendency

Mean or expected value

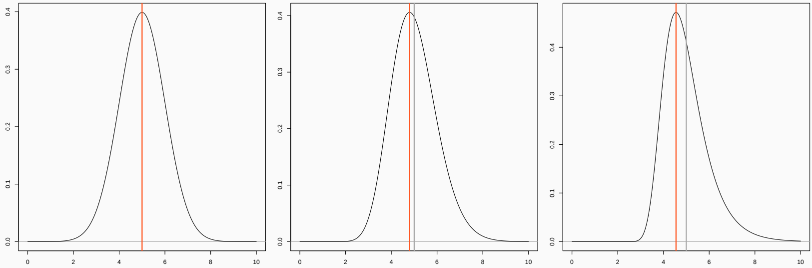


Measures of central tendency

Median

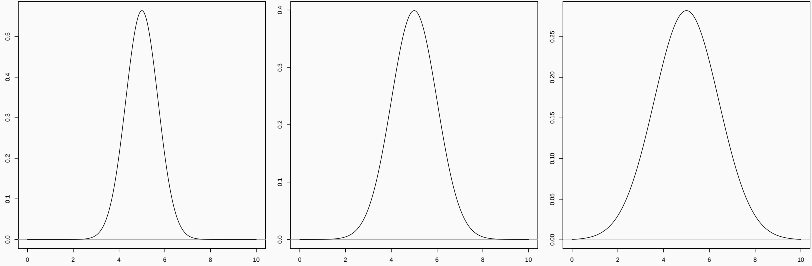


Mode

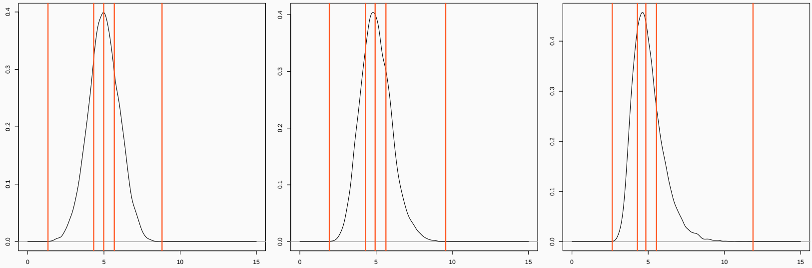


Measures of dispersion

Variance



Order statistics



Statistics

Statistics

What?
The science of collecting and analysing numerical data in large quantities

How?
By planning studies, exploring and modelling the data using the tools of **probability theory**

Why?
To infer properties of a **population** from a **sample**

EXERCISE: probability or statistics?

You have a fair coin. You toss it 100 times. How likely is it to land heads 60 times or more?

Exercise: probability or statistics?

You have a fair coin. You toss it 100 times. How likely is it to land heads 60 times or more?

Probability

- Random process is known (or assumed): ‘fair coin’
  - $\Pr(\text{‘heads’}) = 1/2$
  - $\Pr(\text{‘tails’}) = 1/2$
- Objective: **find probability of a certain outcome**

Exercise: probability or statistics?

I give you a coin. You toss it 100 times and count 60 heads. Is the coin fair?

Exercise: probability or statistics?

I give you a coin. You toss it 100 times and count 60 heads. Is the coin fair?

Statistics

- Outcome is known (or measured): '60/100 heads'
- Objective: **characterise the random process**

Probability theory and statistics

Probability theory

- Defines the model
- …and often its **parameters**

Statistics

- Collects the data
- ‘Fits’ the model (estimates its parameters)
- Makes **inferences**