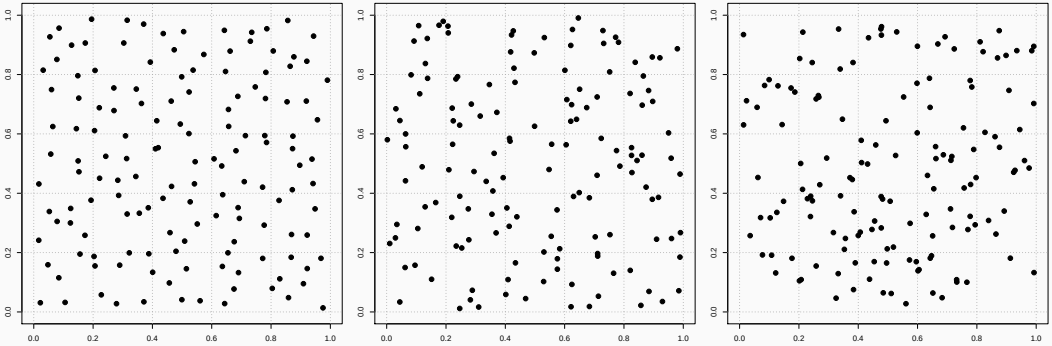
Probability theory and statistics

Gianluca Campanella

Contents

Probability theory

‘Random’ points



What is probability?

'The extent to which something is likely to happen'

— Oxford English Dictionary

Examples

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

Sources of uncertainty

Imperfect information

Current predictive tools can only assign a number indicating our degree of certainty

Stochastic process

The experiment is designed to produce uncertain results (because it's fun)

Probability theory

What?

The branch of mathematics concerned with random phenomena

Probability theory

What?

The branch of mathematics concerned with random phenomena

How?

Using mathematical abstractions of non-deterministic events

Probability theory

What?
The branch of mathematics concerned with random phenomena

How?
Using mathematical abstractions of 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 'average behaviour'

Statistical regularity

In summary…

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

Random variables ‘encapsulate’ random events

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 $X = x$

Example

- X represents the (‘archetypal’) outcome of a coin toss
- $X =$ ‘head’ represents one (actual) outcome
- $\Pr(X =$ ‘head’$)$ is the long-term probability of the outcome ‘head’

Maximum of two fair dice

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

Simplified approximations to reality

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

Characterising probability distributions

Measures of central tendency

- (Arithmetic) mean or average
- Median
- Mode

Characterising probability distributions

Measures of central tendency

- (Arithmetic) mean or average
- Median
- Mode

Measures of dispersion

- Variance
- Minimum and maximum $\rightarrow$ range
- Quantiles (a.k.a. order statistics)

Characterising probability distributions

1, 8, 16, 30, 32, 37, 53, 80, 86, 91, 93

- Mean?
- Median?
- Mode?
- Variance and standard deviation?
- Minimum and maximum, and range?
- Quartiles?

Statistics

Statistics

What?

The science of collecting and analysing numerical data

Statistics

What?

The science of collecting and analysing numerical data

How?

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

Statistics

What?

The science of collecting and analysing numerical data

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

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 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’
- Objective: find probability of a certain outcome

Probability or statistics?

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

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