Intro to the Topic

Part I

Introduction & Basics
Intro to the Topic

Recommended Books for the Course


Assessment

- Exam in May
- Three hours
- Attempt Any *Four* from *Six* Questions

Tutorials

- Tutorial/Workshop every week or so
- One hour long?
- Some questions from tutorials will feature on the exam
A Gartner¹ Definition:

‘Analytics has emerged as a catch-all term for a variety of different business intelligence (BI)- & application-related initiatives . . . Increasingly, “analytics” describes statistical and mathematical data analysis that clusters, segments, scores and predicts what scenarios are most likely to happen.

‘Whatever the use cases, “analytics” has become a hot business topic, gathering interest from business and IT professionals like looking to exploit huge mounds of internally generated & externally available data.’

¹i.e. business-oriented
Mathematical Modelling: What is a Model?

A model is an abstraction or representation of something. Could be:

- An *iconic model* (e.g. airplane in a wind tunnel)
  - represents the system with assumptions,
  - gives a simplified picture of what actually happens,
  - resembles the real thing with observable effects.

- A *mathematical model* (e.g. models of planet undergoing climate change)
  - symbolises some physical (i.e. climate)/conceptual (e.g. budget) subject
  - includes concepts not visually apparent
  - math models the system with math concepts & language.

For some problems, maths are not tractable - must use Probabilistic/Stochastic techniques.

Characteristics of a Model

- highlights features of interest without manifesting the 'unnecessary' detail.
- is cheaper, more convenient, and safer to manipulate than the real-world equivalent.
- is deterministic i.e. produces (on same input) same results on running it, unlike experiment.
Mathematical Modelling: Example 1.1

- **Tablet Dissolution in a digestive system: system features**
  - physically complex (multiple dissolution environments & individual differences)
  - also experimentally expensive (difficult, expensive & non-deterministic)
  - ethically controversial (with human/animal experiments)

  **Model features:** cheap, deterministic & lots of detail

**Maths required:** Differential equations (advection & diffusion)

---

**Figure 1.1:** Drug Dissolution In-Vitro Modelling
Mathematical Modelling: Example 1.2

- Modelling of Option Pricing in Financial Systems
  - need price of a particular financial instrument with time.
  - must solve the Black-Scholes Formula


\[
\frac{\partial V}{\partial t} + \frac{1}{2} \sigma^2 S^2 \frac{\partial^2 V}{\partial S^2} + rS \frac{\partial V}{\partial S} - rV = 0
\]

(a) Black-Scholes Eqn
(b) Option Price Plot

**Figure 1.2:** Option Pricing

Computational Science: What is it?

A Definition from shodor.org

‘the application of computational & numerical techniques to solve large and complex problems. [It] takes advantage of not only improvements in computer hardware, but more importantly, the improvements in algorithms & mathematical techniques.’

Computational Science is the intersection of three disciplines:

**Figure 1.3:** A Schematic of Computational Science shodor.org
Computational Science: What is it (/2)

In Fig. 1.4,

- computational science is a branch of science (application)
- supported by the mathematical methods (algorithms)
- ... and computer science (architecture).

With theory & experiment, a crucial third mode of scientific research & design.

![Diagram](shodor.org)

**FIGURE 1.4: Another View of Computational Science**

13 / 327

Computational Science: What is it? (/3)

In the application of computational science to real-world problems it is necessary to decide the following:

- What scientific event or problem is in focus?
- What are its boundaries & what is external to the system?
- What are the system's parts & what detail should we include?
- Any assumptions to be made about its behavior?
- Have other systems been studied akin to this one?

These decisions made, can put together the math model to represent problem behaviour.
What is a Data Analyst/Data Modeller?

As Data Analyst/Modeller, you look at following data properties:

- Origins? Human/Machine, Single/Multiple Sources?
- Has it been altered/filtered already?
- Is there a lot of it (i.e. for model training/evaluation)?
- What is the granularity?
- How clean is it/Is it noisy?
- What are the units?
- What are its properties?

NB: Data is often someone’s IP and is treated as such by them!

Data Analytics & Developments in Computing

- Computing advances (esp. memory costs & processor speed) have driven Computational Science.
- In particular, we note developments such as:
  - Memory costs decrease with increase in storage capabilities (e.g. Amazon S3)
  - Processing power rise with reduced cost (From $10^{10}$/USD/GFLOP to $10^{-2}$/USD/GFLOP!)
  - Increases in Internet network speeds up to $10^{10}$ bits per s
  - Data availability from multiple sources (smartphones & sensors)

These may be seen as key enablers in Data Analytics.
Data Analytics ‘chews on’ ‘Big Data’. Some facts:\(^3\)

- **Volume**
  - Up to 2003, we created \(5 \times 10^{18}\) Bytes of Data.
  - In 2011, generated the same every two days.
  - In 2013, same created every 10 minutes.
  - In Future???

- **Variety**
  - Types un/structured data (e.g. text, audio, video, -omics).
  - Est.d 2 Bn smartphones in 2015 giving out sensor data.
  - Govt monitors live video feeds from \(5 \times 10^8\) of surveillance cameras.

Add the other V’s: **Velocity & Veracity**...

\(^3\)Source IBM

How do these topics fit in together?

**Figure 1.5:** Big Data Analytics (horicky.blogspot.ie)
Intro to the Topic
Course Introduction
Mathematical Modelling
Computational Science
Data Analytics
Examples of Deep Analytics

Examples of Time Series Data Analysis

Sensor Data Analysis
- Data from sensors on wearer's body (or from smartphone)
- Wearers can be the Young, Seniors or Athletes (or ...)
- Young often have general (health? diet?) data needs
- For Seniors often use-case is memory-related (events)
- For Athletes, purpose is often performance-related.

Figure 1.6: Image Data From Sensors

The Data
- Usually long-run time series (image or sensor data)
- Often quite noisy with gaps (due sensors & network reliability)

The Research Questions
- Is there periodicity present in the data?
- If so, does Granularity Analysis reveal anything?
  - are there 'events' across time scales?
  - what of memory interest could they highlight?
  - do event correlate btw sensors (i.e. image, accelerometer)?

For athletes' sensor data, can ask the questions:
- is the performance of the team as a whole optimal?
- action needed on individual under-performing outliers?
Betting Data Analysis

- Betting data from websites from betting ‘agents’.
- Each ‘agent’ has own strategy (loyalty, motivation) & profile
- Both of these are of interest to betting companies.
- Of equal (greater?) interest is what they as a whole ‘say’.

The Data

- Again multiple long-run time series (odds data)
- One time series for each outcome
- Market varies from liquid (major tennis) to illiquid (low league)
- Good quality, relatively non-noisy data in the main
- Web companies (GAFAs) won’t want to release data (IP!)

Betting Data Analysis (2/2)

The Research Questions for Company

- Identify most valuable customers? (big bettors? risk takers?)
- Need tracking (insider info, responsible betting, ‘cash out’)?
  - What are the major movements in the data (real-time)?
  - Can this be used to optimize ‘cashout’?
  - How about hard-to-price matches (‘Home underdogs’)?
  - Outcomes with unexpected events (e.g. sending off/penalty)?
### Origins of the Module

- **Course began life as a Computational Biology module.**
- It still retains much of that sort of flavour.
- Many problems examined involve time-dependent situations.
- For this reason time series are covered in some depth.
- Many models build on those developed in earlier times.
- An example would be Rumour Spread using a model for Infectious Diseases.

### Maths Content in the Module

Is a knowledge of Higher Level Maths a prerequisite?

- We cover many supplementary topics in the course.
- Topics that you are expected to be familiar with:
  - Basic probability
  - Basic matrix operations
  - Graph plotting
  - Calculus
- There will be tutorials most weeks in the course (3rd hour).
- Tutorial solutions given out towards the end of Semester.
In the CA659 Module we will cover modelling topics such as:

- Time series analysis
- Modelling with Linear & Non-linear Difference Equations
- Modelling with Linear & Non-linear Differential Equations
- Continuous Models of Growth & Decay
- Linear & Non-linear Models of Interaction