

# Probability Theory & Statistics

# Introduction

## Motivation

Data and statistics are the language of modern business and science. It is rare in the modern work place to embark on a task or project that doesn't require some form of collection of data and some form of description of that data. Probability and statistics are *descriptions of data*. Moreover, the language they provide for describing data is rich and nuanced, but requires practice (as with any language be it German or Python!)

## Aim

The sessions' aim is fluency in *statistical language* and *probabilistic thinking*.

## Delivery

The sessions should occur every two weeks, over the period of a year to two years depending on the depth of discussions during sessions.

## Fundamental Objects

There are 4 fundamental objects<sup>1</sup> in probability theory:

1. *Events*
2. *Random Variables*
3. *Distributions*
4. *Numbers*

We will work through and discuss each of these objects and most importantly *practice using them in problems*.

## Structure

### *Prerequisites*

Each session tries to be self-contained, although this is sometimes difficult or impossible for certain topics, in which case a *prerequisite* list of sessions or reading is suggested at the start of the session document. Further reading is also provided at the end of each session for a more in-depth discussion of the session topic.

### *Notation*

New terms and concepts are *emphasized*.

---

<sup>1</sup>Thinking in terms of objects avoids *category errors*. The most common source of confusion when discussing data and its descriptions.

### ***Interpretation***

We will try to avoid too much mathematical jargon and unnecessary proofs, keeping formal notation to the bare minimum and where necessary explained in plain language.

### ***Discussions***

Frequently asked questions from each session are added to the end of the session notes for future reference.

### ***Practice***

Practice questions and their solutions are provided, the expectation is that they are worked through during the session as an informal discussion, rather than set as exercises.

### **Sessions**

List of the sessions in suggested reading order

1. *Events*
2. *Counting*
3. *Conditioning*
4. *Distributions*
5. *Inequalities*
6. *Simulation*
7. *Intervals*