Statistics for Data Science

Lecture 1

Introduction

Ken Cameron

Statistics for Data Science

- Lecturers
- Content
- Assessment
- Structure
- Resources

Lecturers

- Vangelis
    - Evangelos Evangelou
    - Dept. of Mathematical Sciences
    - E.Evangelou@bath.ac.uk
    - 4 West 4.7

- Ken Cameron
    - Dept. of Computer Science
    - K.M.Cameron@bath.ac.uk
    - 1 West 3.53

Content (My half)

- Sets
- Laws of Probability
- Random Variables
- Bayes Theorem
- Discrete & Continuous Distributions
- Multiple Random Variables
- Sampling
- Central Limit Theorem

Assessment

- Examination
  - 50% of overall unit mark.
  - 2 hours, during exam period.
- Class Tests
  - 50% of overall unit mark.
  - In classroom, one mid semester, one at the end.
  - 30 minute written.

Structure

- Lectures
  - Tuesdays 11.15 am and Thursdays 1.15pm
- Tutorials
  - Second slot on Thursdays 2.15pm
  - Will cover material from preceding week.
  - Problem sheets.

Resources

- Lecture slides/notes on moodle.
- Problem sheets on moodle.

- Statistical Inference. Casella, Berger.

Statistics

- Two of the most commonly used statistics that we can calculate from some data points x_i

- Arithmetic mean

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

- Standard deviation

$$\sigma = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \mu)^2}$$

Probability

- A probability is a number between 0 and 1 which measures the chance of an event occurring

- Examples
  - The chance of throwing a 6 on a fair die
  - The chance that a student is between 1.7-1.8m
  - The chance that it will rain tomorrow, given that it is sunny today

Probability

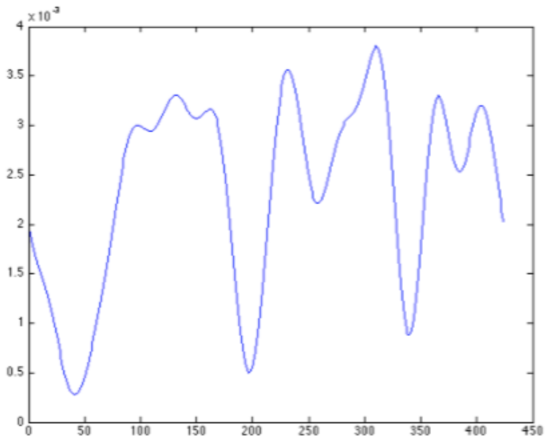
- In practice, you will often just count the occurrences of events and divide by the total
  - e.g. (number of rainy days)/(total days in year)

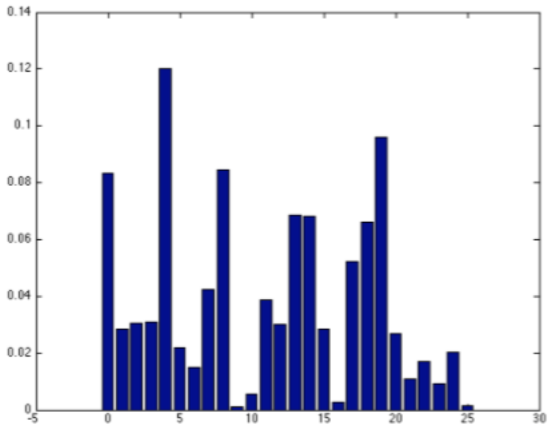
- You might have reason to expect that next year will be wetter than this year.

- Frequentist: stick with counting
- Bayesian: incorporate our prior beliefs

Probability

- Probabilities can be continuous or discrete





- Continuous: height, computation time
- Discrete: coin toss, die roll, horses in race

Probability Notation

- A *random variable* is a variable whose value is subject to uncertainty or chance

- It can take on one of many values, called *random variates*

- For example, let x be a random variable representing the roll of a die.

  - Then the set of outcomes is $\mathbb{X} = \{1,2,3,4,5,6\}$, and we call each element $X \in \mathbb{X}$ a random variate

Probability Notation

- Write $p(x = X)$ to mean the probability that the random variable x takes on the value X
  - e.g. for a fair die $p(x = 3) = {}^{1}/_{6}$
- We might write other conditions on x
  - e.g. $p(x < 4) = 0.5$
- We will sometimes write $p(x)$ when the meaning is clear
- We might also write $p(3)$

Joint Probability

- The ***joint*** probability is the chance that a collection of events occur together

- We write

$$p(x = X, y = Y, z = Z)$$

- Or simply

$$p(x, y, z)$$

- Exchangeability

$$p(x, y, z) = p(x, z, y) = p(y, z, x) = \cdots$$

Marginal Probability

- The ***marginal*** probability is the chance of observing a random variable in a particular state when two (or more) events are observed simultaneously

- Given variables x and y, the joint is $p(x, y)$
  the marginal is:

$$p(x) = \sum_{y} p(x, y)$$

Conditional Probability

- The ***conditional*** probability is the chance of observing an event $y$given that an event $x$has occurred

- This is written $p(y|x)$, and can be calculated:

$$p(y|x) = \frac{p(x,y)}{p(x)}$$

Example

- We have two random variables, the type of newspaper and the day of the week

	Voice	Opinion	Record
Mon	10	8	9
Tue	8	6	7
Wed	7	8	5
Thurs	8	7	7
Fri	9	8	8

Table 2.1: Newspaper Sales in 10,000's

- What are the variates?

- What is:

  $p(\text{Voice}, \text{Monday})$? *Joint*

- What is:

  $p(\text{Monday})$?

  Marginal

- What is:

  $p(\text{Voice}|\text{Monday})$?

  Conditional

Fundamental Axioms

1. All probabilities are non-negative, for all x
$$p(x) \geq 0$$

2. Mutually exclusive events (cannot occur at the same time) add
$$p(x_1 \text{ xor } x_2 \text{ xor } \dots) = p(x_1) + p(x_2) + \cdots$$

3. Some event must occur
$$\sum_x p(x) = 1$$

Fair Dice Roll Example

1. All outcomes (1-6) have positive probability
2. Probabilities of rolling a 1 or a 2 add $p(1 \text{ or } 2) = p(1) + p(2) = {}^{1}/_{6} + {}^{1}/_{6} = {}^{1}/_{3}$
3. We are guaranteed to get some result 1-6 $p(1 \sim 6) = 6\ \times {}^{1}/_{6} = 1$

- If you ever find your probabilities are negative or do not add up to one (over all events), something has gone wrong

Set Theory

- We can use Set Theory to give us a more formal basis on which to build.

- The set S, of all possible outcomes of an experiment is called the sample space.

- An event is any collection of possible outcomes of an experiment.
  - i.e. Any subset of S.
  - Let A, a subset of S, be an event.
  - We say A occurs if the outcome of the experiment is in the set A.

Set Theory

$$A \subset B \Leftrightarrow x \in A \Rightarrow x \in B$$

$$A = B \Leftrightarrow A \subset B \text{ and } B \subset A$$

- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$

- Complementation: $A^c = \{ x : x \notin A\}$

Sigma Algebra

- A collection of subset of S is called a sigma algebra or Borel field, B if it satisfies the following properties:
  - $\emptyset \in B$ (the empty set is an element of B)
  - If A $\in B$, then $A^c \in B$ (B is closed under complementation)
  - If $A_1, A_2, \ldots \in B$, then $\cup_{i=1,\infty} A_i \in B$ (B is closed under countable unions)

Probability Function

- Given a sample space S and an associated sigma algebra B, a probability function is a function P with domain B that satisfies
  - $P(A) >= 0$ for all $A \in B$
  - $P(S) = 1$
  - If $A_1, A_2, \ldots \in B$ *are pairwise disjoint, then* $P(\cup_{i=1,\infty} A_i) = \sum_{i=1,\infty} P(A_i)$

A Theorem

- If P is a probability function and A and B are any sets in $\mathcal{B}$ then,

  a. $P(B \cap A^c) = P(B) - P(A \cap B)$
  b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
  c. If $A \subset B$, then $P(A) <= P(B)$

- Can we prove this?

a. $P(B \cap A^c) = P(B) - P(A \cap B)$

- For any sets A and B:

$$B = \{B \cap A\} \cup \{B \cap A^c\}$$

And therefore

$$P(B) = P(\{B \cap A\} \cup \{B \cap A^c\}) = P(B \cap A) + P(B \cap A^c)$$

Remember $B \cap A$ and $B \cap A^c$ are disjoint. Rearrange to get a.

b. $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

- Start with the identity:

$$A \cup B = A \cup \{ B \cap A^c\}$$

$B \cap A$ and $B \cap A^c$ are disjoint

$$P(A \cup B) = P(A) + P(B \cap A^c) = P(A) + P(B) - P(A \cap B)$$

c. If $A \subset B$, then $P(A) <= P(B)$

- I’ll leave you to consider...

- Answer next lecture.