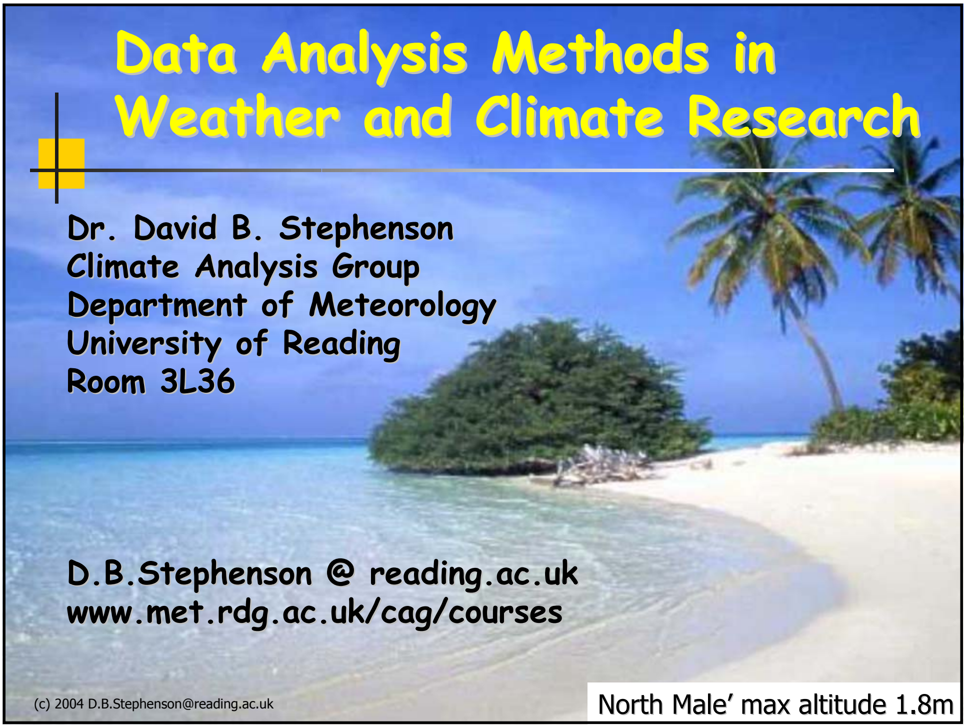


Data Analysis Methods in Weather and Climate Research

Dr. David B. Stephenson
Climate Analysis Group
Department of Meteorology
University of Reading
Room 3L36

D.B.Stephenson @ reading.ac.uk
www.met.rdg.ac.uk/cag/courses

North Male' max altitude 1.8m

Course outline

1. Introduction
2. Descriptive sample statistics
3. Basic probability concepts
4. Probability distributions
5. Parameter estimation
6. Statistical hypothesis testing
7. Basic linear regression
8. More advanced regression
9. Introduction to time series

3. Basic Probability concepts

1. "Events" and "event space"
2. Definitions of "probability" and odds
3. Joint and conditional probability
4. Random variables
5. Expectation, covariance, and correlation

Motivation

Probability is THE concept needed to make the link between a sample of data and the whole population. It is also the key idea for understanding and modelling uncertainty and risk.

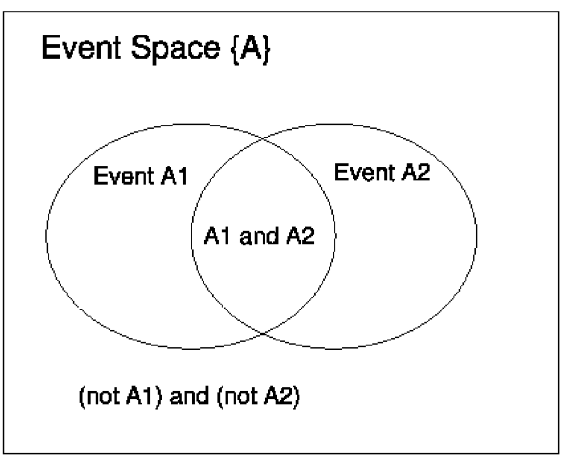
3.1 "Events" and "event space"

An event A is a set of possible outcomes of an uncertain process e.g. A={Heads}, {Tails}, {Wet day}, {T>20C}, etc.

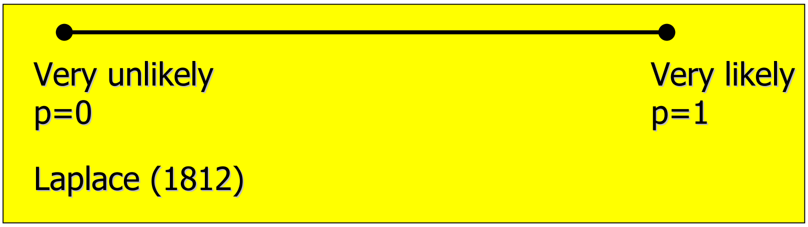
Events can be simple (indivisible) or compound

Event space is the set of all possible events.

Euler (Venn) diagram



3.2 Definitions of “probability”



1. Number of symmetric ways
2. Relative frequency of an event repeated many times
3. Non-frequentist subjective definition
4. Axiomatic definition

3.2a Number of symmetric ways

Pr{A} = number of ways A can occur divided by the total number of ways

e.g. Pr{odd number on die}=3/6 since 3 ways to get odd number {1,3,5} out of a total of 6 equally likely outcomes {1,2,3,4,5,6}

3.2b Relative frequency

Suppose an event occurs m times in n repeated trials, then the relative frequency m/n provides an estimate of the probability in the limit as n→infinity.

This is the so-called "frequentist" approach based on the "Law of large numbers".

3.2c The subjective approach

Probability = price of a <u>fair</u> bet divided by the amount you could win.

Mean profit=p(W-B)+(1-p)(-B)=0

Odds = Prob{event}/Prob{not event}=p/(1-p)

3.2d The axiomatic approach

- All probabilities Pr{A} >= 0
- The probabilities of all events sum to one
- Pr{A or B}=Pr{A}+Pr{B} if A and B are mutually exclusive events

3.3 Joint and conditional probability

Joint probability = Pr{A and B}

Conditional probability = Pr{A and B}/Pr{B} = Pr{A|B}
i.e. probability of event A GIVEN that event B occurs

Bayes' theorem: Pr{A|B}=Pr{B|A}Pr{A}/Pr{B}
Useful for getting from prior Pr{A} to posterior Pr{A|B}
e.g. conditioning hypothesis A based on new data B

3.4 Random variables

Random variable (r.v) = a number X associated with a random event.

- Categorical e.g. X=0 (dry) or 1 (wet day)
- Discrete count variable: X=0,1,2,3,…
- Continuous variable: X=real number e.g. temperature, rainfall amount, etc.

Note: observed values of X are denoted x.

3.5 Expectation (population mean)

$$E(X) = \sum_{j=1}^{m} X_j \Pr\{X = X_j\}$$

Sometimes referred to as the population mean μ_X

Not to be confused with the sample mean $\overline{x}$

3.5 Covariance

$$\text{var}(X,Y) = E((X - E(X))(Y - E(Y)))$$

Properties:

- When X=Y gives population variance $\text{var}(X,X) = \sigma_X^2$
- Independent of means of X and Y

3.5 Correlation

$$cor(X,Y) = \frac{\text{var}(X,Y)}{\sqrt{\text{var}(X)\,\text{var}(Y)}}$$

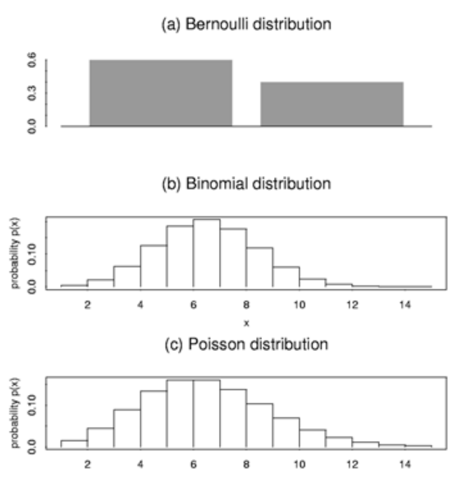
Properties:

- Measure of linear association varying between –1 and +1
- Independent of mean and variance of X and Y
- Symmetric in X and Y

4. Probability distributions

1. Continuous vs. discrete distributions
2. Discrete distributions
   1. Bernoulli distribution X~Be(π)
   2. Binomial distribution X~Bin(n,pi)
   3. Poisson distribution X~Poisson(mu)
3. Continuous distributions
   1. Uniform distribution X~U(a,b)
   2. Normal (Gaussian) X~N(mu,sigma)
   3. Gamma distribution X~Gamma(alpha,beta)

Discrete distributions



Continuous distributions