

Probability and Statistics

This course will teach and understanding of probability and the skills needed to analyse data.

Overview

Computer science is the science of computation and data. Our approach to data is typically very sophisticated and algorithm driven, however, it is often different from the approaches used in many of the other disciplines to which computer scientists contribute. In many areas of science and in enterprise, data is analysed using a frequentist approach which has been developed over the last two centuries into a powerful and practical method for understanding and interpreting data, particularly the data that results from experiments, whether the empirical results of scientific experiments or data collected in pursuit of business goals. The aim of this course is to introduce computer science students to these methods and to teach them statistical skills which are both useful as an approach to data and as the language of data commonly employed in science and industry.

Aims

The aim of this course is to introduce students to mathematics underpinning statistics, to the methods commonly used to analyse data.

Learning outcomes

At the end of this course the students will:

- Understand the foundations of probability and statistics.
- Be familiar with sampling and sampling bias.
- Know and be able interpret different distributions.
- Make the best use of descriptive statistics.
- Be able to analyse data using classical statistical tests.
- Understand the design and analyse of experiments.

Syllabus

- Random variables, probabilities, probability mass functions and Bayes rule.
- Counting and selection; the binomial distribution.
- Moments of a distribution include the mean, variance; sampling and bias.
- Poisson processes, the Laplace and gamma distribution.
- The law of large numbers. The central limit theorem: the Gaussian distribution.
- Simulating samples from specified distributions.
- Hypothesis testing. The Student t-test.
- Test of normality.
- Summary statistics: the Wilcoxon signed rank test and Wilcoxon-Mann-Whitney U test.
- Regression, slope and intercept, the statistical significance of regression.
- Understanding basic design principles, one-way ANOVA, the Kruskal-Wallis and Friedman tests.

Teaching and assessment

The course will include lectures and practical workshops, the workshops will include marked coursework worth 10% of the final mark. A 2 hour exam will make up the remaining 90%.