


FOUNDATIONS OF STATISTICAL INFERENCE


Technical Teaching
STATS
KANAGAWA Motonobu

ABSTRACT

Statistics is a foundation of many areas of science and engineering that involve “data.” This course focuses on fundamental concepts in statistical inference that are necessary for applying statistical methods in practice and that form the basis of other fields such as machine learning.

Teaching and Learning Methods: Students learn by lectures, exercises, and computer experiments.

Course Policies: The valid usage of statistical methods requires a mathematical understanding of the underlying mechanism. As such, the course covers both mathematical and algorithmic aspects of statistics.

BIBLIOGRAPHY

1. B. Efron and T. Hastie, “Computer Age Statistical Inference: Algorithms, Evidence and Data Science”, Cambridge University Press, 2016. The textbook is freely available at the authors’ website: <https://web.stanford.edu/~hastie/CASI/index.html>
2. J. O. Berger, “Statistical Decision Theory and Bayesian Analysis”, Springer, 1985.

REQUIREMENTS

The language of statistics is probability theory. As such, the course requires basic knowledge of probabilities in finite discrete outcomes, that of calculus (such as differentiation and integration), and that of linear algebra (such as eigenvectors and eigenvalues of a matrix, and the solution of linear equations).

DESCRIPTION

The course focuses on fundamental concepts in statistics, using the simplest examples such as estimation of the mean from a finite sample and linear regression. It covers both frequentist and Bayesian approaches. The former includes statistical hypothesis testing and maximum likelihood estimation. The latter includes the notion of prior and posterior distributions and Bayes’ rule. I will explain how these two approaches differ in interpreting the “data” and “model”, and in defining “optimal” decisions. I will also teach Monte Carlo methods, a key ingredient in both approaches.

Learning outcomes: Students will learn fundamental notions of statistics and the way of statistical thinking. These include: 1) Conclusions and decisions that can be made from data are highly dependent on how the data are obtained. The key concepts in this regard are selection bias and consistency. 2) There are various trade-offs in statistical methods, notably those between i) bias, ii) variance and iii) computation. This knowledge is necessary in practice, as it is required for the selection of a method and a model, and hyper-parameter tuning. The key concepts here include the generalization error, cross-validation, regularization and the curse of dimensionality.

Nb hours: 21 hours

Nb hours per week: 1,5 hours

Grading Policy: Exam and report.