30470303-0: Probability and Statistics

Yuhao Wang

Fall Semester, 2020-2021

Course description: Statistical science provides powerful tools to help people extract useful information from massively collected data. Since its first establishment by Karl Pearson in the late 1800s, statistical science has created a huge impact on social science, economics, natural science, computer science, advanced manufacturing etc. across the entire globe. With the advent of information technology and the explosion of large scale data, statistical science is becoming increasingly important in modern society. This is also the main purpose to open this class to IIS students.

In this course, we provide a comprehensive study of the development of statistical science in the last 100 years, from its probabilistic foundations to modern statistical methods used in high-dimensional and nonparametric systems. The end goal of this course is to equip students with the ability to read most recent statistical papers published in top statistical journals and to apply advanced statistical techniques to their research or industrial tasks.

Main contents: This course covers the following five broad topics:

- Basic background in probability

  - definitions of probability, law of large number, central limit theorem, basic concentration inequalities and tail bounds, regression analysis ..

- High-dimensional estimation

  - high-dimensional sparse regression, matrix estimation, high-dimensional inference ..

- Asymptotic statistics and confidence intervals

  - Bootstrap, semi-parametric models, hypothesis testing ..

- Minimax lower bounds

  - lower bound and hypothesis testing, minimax lower bounds in high-dimensional / nonparametric models ..

- Statistics and machine learning

    - conformal prediction ..

Instructors:

Yuhao Wang, Tsinghua FIT Building 4-6018,
email: yuhaow@tsinghua.edu.cn,   Web: https://yuhaow.github.io

Teaching assistants:

Tong Zhang, Tsinghua FIT Building 4-600-063,
email: zhangton20@mails.tsinghua.edu.cn

Zhiheng Zhang, Tsinghua FIT Building 4-600-068,
email: zhiheng-20@mails.tsinghua.edu.cn

Grading: 30% homework, 30% mid term exam, 35% final term exam, 5% scribing.

- Homework: Problem sets will be released approximately each 2-3 weeks;

- Scribe: Each student should scribe at least one class in this course.

Time and place: Lectures are Thursday 9:50 - 12:15, 6A201; Office hours are TBD.

Books and References: This lecture is constructed based on a variety of textbooks and research papers. The main books we refer to are listed in the following:

M. J. Wainwright. *High-dimensional Statistics: A Non-asymptotic Viewpoint.* Cambridge University Press. (**Main textbook**)

R. Vershynin. *High-dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press.

R. Durrett. *Probability: Theory and Examples.* Cambridge University Press.

L. Wasserman *All of Statistics: A Concise Course in Statistical Inference.* Springer Science & Business Media.

A. W. Van der Vaart. *Asymptotic statistics.* Cambridge University Press.