

ORF 524

Fall Semester, 2004

TTh 3:00 – 4:20 in Benheim Finance Center

Home page: <http://www.orfe.princeton.edu/~jqfan>

Instructor: Jianqing Fan, Professor. Office: E-332 E-quad. Phone: 258-7924. E-mail: jqfan@princeton.edu

Office Hours: Tuesday 11:00-12:00 (E-332, E-quad), Thursday (4:30–5:30, 102 BCF), or by appointments.

Textbook: Bickel, Peter J. and Doksum, Kjell A. (2001), *Mathematical statistics : basic ideas and selected topics*, 2nd ed. Upper Saddle River, NJ: Prentice Hall.

Reference:

- Rice, J.A. (1995). *Mathematical Statistics and Data Analysis*. (2nd ed.), Duxbury Press.
- Casella, George and Berger, Roger L. (2002), *Statistical inference*, 2nd ed., Pacific Grove, CA: Duxbury
- Lehmann, Erich L. and Casella, George (1998), *Theory of Point Estimation*, 2nd ed., New York: Springer
- Lehmann, Erich L. (1997). *Testing statistical hypotheses*, 2nd ed., New York, Springer

Syllabus: Course material will be covered the following chapters (in order):

- Chapter 1: Statistical Models (1. Models, parameters, and statistics; 2. Bayesian models; 3. Sufficiency; 4. Exponential families)
- Chapter 2: Methods of Estimation (1. Heuristic method; 2. Estimation equation; 3. Maximum likelihood; 4. E-M Algorithm)
- Chapter 3: Measure of Performance (1. Decision theoretic framework 2. Bayesian procedures; 3. Minimax procedures; 4. Unbiased estimation and risk inequalities)
- Chapter 4. Testing and confidence regions (1. Introduction 2. Neyman-pearson's lemma; 3. UMP tests; 4. Confidence intervals; 5. Duality between confidence regions and tests; 6. Prediction intervals; 7. Likelihood ratio procedures)
- Chapter 5 (combined with Chap 6 of the book) Asymptotic Approximations (1. Meaning and uses of asymptotics; 2. Consistency; 3. Delta method; 4. Asymptotic estimation theory. 5. Large sample tests and confidence regions. 6. Asymptotics of posterior distribution)
- Chapter 6. Linear Models and Generalized Linear Models (1. Least-squares and related quantities; 2. Inference about linear models; 3. Prediction; 4. Introduction to GLIM)

Homework: Problems will be assigned at class meetings and will be due in class on Tuesdays of following weeks. No late homework will be accepted. Missed homework will receive a grade of zero. The lowest score will be dropped. The homework will be graded, and each assignment carries equal weight. You are allowed to work with other students on the homework problems, however, verbatim copying of homework is absolutely *forbidden*. Therefore each student must ultimately produce his or her own homework to be handed in and graded.

Exams: There will be one in-class midterm exam, and a final exam. All exams are required and there will be no make-up exams. Missed exams will receive a grade of zero. All exams are open-book and open-notes. Calculators may be used during the exams.

Schedules and Grading Policy:

Homework (25%) due in class every Wednesday
Midterm Exam for *this class* (25%) Tuesday, November 16, 2004 (tentative).
FINAL EXAM for *this class* (50%) January 13, 2005 (tentative).

Introduction

Exponential increases in computing power and falling costs have had a profound impact on the development of statistics and applied mathematics. The field of statistics has experienced extraordinary growth in the last two decades. Many data-analytic techniques have been developed and many new phenomena have been unveiled. They have become indispensable tools in contemporary scientific research, technological invention, knowledge discovery, and policy making. With such exciting developments and advancements, statistics have played important and critical roles in many scientific disciplines such as genomics, biomedical studies, financial economics, machine learning and information engineering.

Technological invention and the information age have revolutionized scientific research and technological development. Modern technologies allow scientists to collect data in unprecedented size and complexity. Examples include microarrays, proteomics, fMRI brain images, functional data and longitudinal data, financial and business activities, among others. Quantitative methods have been widely employed in different scientific and engineering disciplines, which empowers knowledge discovery and policy making. Successful examples have been scattered around the statistical and applied mathematical literature. Yet, many advanced techniques have not been fully developed and employed; many complex data sets have not been fully explored and analyzed; many new business and new products pose new and complexity statistical problems. These give statistics opportunities with challenges.

Depending on the background of applications, data can be collected of following types:

1. Continuous measurements (e.g. income, weights, blood pressure)
2. Discrete $\left\{ \begin{array}{l} \text{Categorical (e.g colors, gender, preferences, ranking)} \\ \text{Numerical (counts)} \end{array} \right.$

Individual observations can be

1. Scalars: Expenditures, measurements.
2. Vectors: Responses along with covariates (e.g. income, education, gender, age, experience)
3. Longitudinal data (panel data): Individual characteristics and responses are collected at subject-specific times.
4. Functional data: growth curve, movement curves, signals of car crashes.
5. Images: e.g. Cornea maps, proteomic data, microarray data.

Even though the background of application differs substantially, statistical idea and principles remain the same. This class intends to

- to provide basic tools of mathematical statistics;
- to learn fundamental principles in statistics;
- to understand the basic concepts on the foundation of statistics;
- to deal with a variety of statistical problems and models.

With these tools and principles, one should be able to invent new statistical techniques for his/her own problems.