



Columbia University
MAILMAN SCHOOL OF
PUBLIC HEALTH
DEPARTMENT OF BIostatISTICS

P9111: Asymptotic Statistics

CLASS SESSION

Fridays 1:00-3:30 pm. Location: Biostatistics Computer Lab, Room 656, MSPH.

INSTRUCTOR

Professor Ian W. McKeague, Ph.D.

Office: 722 West 168th Street, Room 639

Phone: (212) 342-1242

e-mail: im2131 "at" Columbia "dot" edu

webpage: <http://www.columbia.edu/~im2131>

Office Hours: Fri 3:30-4:30, or by appointment on Thursday afternoon

COURSE DESCRIPTION

This course will provide a comprehensive introduction to the field of asymptotic statistics. The treatment will be both practical and mathematically rigorous. The course will consist of two parts. The first will be a review of most of the standard topics of limit theory, such as the delta method and central limit theorems, while avoiding many technicalities. The second will present advanced topics such as semiparametric models, empirical likelihood, the bootstrap, and empirical processes. These powerful research techniques are becoming increasingly important for the development of biostatistical methods to handle complex data sets. The overall goal of the course is to train students in the use of advanced asymptotic techniques for medical and public health applications. This course is intended for second-year Biostatistics Ph.D. students to provide a review of asymptotic statistics for the Ph.D. qualifying exam, and give them exposure to a variety of advanced topics.

PREREQUISITES

Theory of Statistical Inference I and II (P9109 and P9110) are prerequisites for most students, but this requirement may be waived with the permission of the instructor. Some familiarity with Real Analysis is essential; Measure Theory is not required, but would be helpful.

COURSE LEARNING OBJECTIVES

Students who successfully complete this course will be able to

- Use asymptotic methods to address problems in statistical modeling that arise from medicine and public health;
- Gain access to the current research literature in mathematical statistics;
- Solve problems on asymptotic methods that appear in the Ph.D. qualifying exam;

- Establish asymptotic results concerning estimators and test statistics in a variety of complex models.

ASSESSMENT AND GRADING POLICY

- 4 homework assignments (40%) graded individually (10% each)
- In-class presentation (20%)
- Final exam (40%).

The final course grades will follow the Mailman School grading policy where:

A+ = 98-100 (Highly Exceptional Achievement)

A = 94-97 (Excellent. Outstanding Achievement)

A- = 90-93 (Excellent work, close to outstanding)

B+ = 88-89 (Very good. Solid achievement expected of most graduate students)

B = 84-87 (Good. Acceptable achievement)

B- = 80-83 (Acceptable achievement, but below what is generally expected)

Homework (40% of Final Grade):

Homework assignments will be announced in class, and will consist of selected problems related to the material covered in lectures or in class readings (see page 4). Solutions to homework problems will be discussed in class.

Guidelines for Homework:

1. You are strongly encouraged to typeset your homework in Latex. Help will be given in class on getting started with Latex.
2. Your solutions should be coherent and grammatically correct. Use simple, clear and complete sentences. Use paragraphs to separate different steps in your solutions.
3. Mathematics is prose! Each statement should be a sentence, with a subject, object and verb. End an equation with a period if it is at the end of a sentence. An = sign can operate as a verb. If at all possible, avoid starting a sentence with notation.
4. Don't use unnecessary words. Notation should be introduced to cut down on tedious repetition of long expressions.
5. Use scratch paper to do exploratory work. If you are asked to prove something for all finite n , special cases (e.g., $n = 1$, $n = 2$) are considered exploratory. Note: Induction can only prove a statement for finite n , and the case $n = 1$ will require a separate argument.
6. Good Samaritan Rule: when you need to use a standard result, mention its *name*, and not a theorem number. If the result has no name, then you should state it (and include a proof if it is not a standard result from class or from real analysis). Don't assume the reader knows what you are about to do—it is usually helpful to outline the overall strategy of the solution, or even each step, before plunging into details.
7. Write out the question before giving the solution. Answer the problems in the order in which they were assigned. Staple the sheets of paper together.
8. If you introduce some non-standard notation that was not specified in the problem, you need to define it. A common omission is to use an ϵ implicitly, without previously noting "Fix $\epsilon > 0$ " say.

9. Your work will generally be more readable if you use displayed equations rather than embedding long equations in the text.
10. Each step of your solution needs to be justified, either by naming a standard result, or filling in the gap by a separate argument. If you are unable to fill the gap (or do any part of the problem), say so!

In-class Presentation (20% of Final Grade):

Each student will select a recent journal article (with the approval of the instructor) to present in class. The article should be related to the material covered in the second part of the course.

Final Exam (40% of Final Grade):

The final exam for the class (3.5 hrs) has two parts: (1) a 3-hour closed-book exam consisting of 7 problems of the type covered in the homework assignments, (2) a half-hour oral exam. The written part of the final exam will emphasize material covered in the first part of the course.

COURSE REQUIREMENTS

Required Textbook: van der Vaart, A. *Asymptotic Statistics*. Cambridge University Press, 2000.

Class participation: You are expected to participate in class discussions by making thoughtful and informed comments, and by asking pertinent questions about the material being covered. You are expected to be responsive to questions asked by the instructor.

Homework: Late homework will not be accepted unless prior permission has been given. You are encouraged to discuss homework problems with fellow students, but your solutions must be your own work. If you obtain any help with a homework problem, you are required to acknowledge the source of the help.

Attendance: If you are unable to attend a class, an explanation must be provided to the instructor.

Missing the final exam: rescheduling will be considered on a case-by-case basis.

COURSE STRUCTURE

The course will be split into two parts: a review of the basic techniques of limit theory (4 or 5 weeks), and advanced topics (8 or 9 weeks).

Limit theory: Types of convergence, continuous mapping theorem, Slutsky's theorem, delta method. Laws of large numbers, Glivenko–Cantelli theorem, central limit theorems.

Advanced topics: M-estimation, Z-estimation, weak convergence of stochastic processes. Empirical processes. Brownian motion, Brownian bridge, martingale and counting process methods. Efficiency of estimators and tests, non- and semiparametric methods in survival analysis, bootstrap methods, empirical likelihood.

COURSE SCHEDULE (LECTURE TOPIC BY WEEK)

1. Why asymptotic statistics? They enable us to find approximate tests and confidence regions, and they can be used theoretically to study the quality of statistical procedures.
2. A review of the basic modes of convergence of sequences of stochastic vectors, along with key techniques such as the Borel–Cantelli lemmas, the continuous mapping theorem and Slutsky’s theorem. Quantile functions and the Skorohod representation.
3. Laws of large numbers, characteristic functions and central limit theorems.
4. The delta method consists of using a Taylor expansion to approximate a function of a random variable (r.v.) by a polynomial in the r.v. This simple but useful method allows us to deduce limit laws in many statistical settings. The method will be illustrated by an application to variance stabilizing transformations.
5. Introduction to empirical process theory. The empirical distribution of a random sample is the uniform discrete measure on the observations. Convergence of this measure leads to laws of large numbers and central limit theorems that are uniform in certain classes of functions.
6. Weak convergence of stochastic processes, the functional delta method. Entropy for Glivenko–Cantelli and Donsker Theorems. Bootstrapping empirical processes. Brownian motion, Brownian bridge.
- 7.-12. Possible choices of advanced topics:
 - M-estimators and empirical processes. The argmax continuous mapping theorem.
 - Contiguity, quadratic mean differentiability, local asymptotic normality.
 - Martingale and counting process methods in survival analysis. Kaplan–Meier and Nelson–Aalen estimators. Cox proportional hazards model.
 - Empirical likelihood. A nonparametric method of inference based on a data-driven likelihood ratio function; like the bootstrap, does not require the user to specify a family of distributions for the data.
 - Overview of semiparametric inference and efficiency. Score functions, estimating equations, tangent spaces.
13. In-class presentations by students.

READING (BY WEEK)

1. van der Vaart Chapter 2
2. Shorack Chapters 2 and 7
3. Shorack Chapters 10, 11, 13, 14; van der Vaart and Wellner 1.1–1.10
4. van der Vaart Chapter 3
5. van der Vaart Chapter 5
6. van der Vaart Chapter 18, 19; Kosorok Chapter I.2
- 7.-12. (depending on the choice of advanced topics) van der Vaart Chapters 6, 7, 20, 25; Kosorok Chapter II; van der Vaart and Wellner Chapters 2 and 3; Kosorok Chapters I.3 and I.4; Martinussen and Scheike Chapters 2-5; ABGK Chapter II; ABGK Chapters IV and VII; Owen Chapters 1-3, 6 and 8; Lehmann and Romano Chapters 12 and 13.
13. Recent journal articles selected by students.

Supplementary references:

Shorack GR, *Probability for Statisticians*

Lehmann, EL and Romano JP, *Testing Statistical Hypotheses*. Third Edition.
van der Vaart A and Wellner JA, *Weak Convergence and Empirical Processes*
Kosorok M, *Introduction to Empirical Processes and Semiparametric Inference*
Martinussen T, Scheike TH, *Dynamic Regression Models for Survival Data*.
Andersen PK, Borgan O, Gill RD, Keiding N, *Statistical Models based
on Counting Processes*.
Owen A, *Empirical Likelihood*.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath.
<http://mailman.columbia.edu/about-us/school-mission/>

Academic Integrity

Students are required to adhere to the Mailman School Honor Code, available online at
<http://mailman.columbia.edu/honorcode>.

The GSAS guidelines on Academic Integrity and Responsible Conduct of Research are available online at
<http://www.columbia.edu/cu/gsas/sub/bulletin/policies/conduct/integrity.html>.

“Scholars draw inspiration from the work done by other scholars; they argue their claims with reference to others’ work; they extract evidence from the world or from earlier scholarly works. When a student engages in these activities, it is vital to credit properly the source of his or her claims or evidence. To fail to do so would violate one’s scholarly responsibility.”

“Plagiarism includes obtaining all or part of a paper (including obtaining or posting a paper online), copying from or paraphrasing another source without proper citation or falsification of citations; and building on the ideas of another without citation. Graduate students are responsible for proper citation and paraphrasing, and must also take special care to avoid even accidental plagiarism. The best strategy is to use great caution in the handling of ideas and prose passages: take notes carefully and clearly mark words and ideas not one’s own. Failure to observe these rules of conduct will result in serious academic consequences, which can include dismissal from the university.”

Disability Access

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu. If you have already registered with ODS, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Lillian Morales (lm31@columbia.edu), the School’s liaison to the Office of Disability Services.