Columbia University
MAILMAN SCHOOL OF
PUBLIC HEALTH
DEPARTMENT OF BIOSTATISTICS

P9111: Asymptotic Statistics

Class Session

Fridays 1:00-3:50 pm. Location: 6th Floor Conference Room, MSPH

Instructor

Dr. Ian McKeague
Office: R639, 722 West 168th Street
Phone: (212) 342-1242
e-mail: im2131xy where x=@ and y=columbia.edu
webpage: http://www.columbia.edu/~im2131
Office Hours: Fri 4:00-5:00, or by appointment

Course Description

This course is intended for second-year Biostatistics Ph.D. students, providing exposure to a variety of advanced topics in asymptotic statistics. The treatment will be both practical and mathematically rigorous. After a brief review of limit theory that was covered in Statistical Inference I and II, we will move on to 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.

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 skills to access the current research literature in mathematical statistics;
- Establish asymptotic results concerning estimators and test statistics in a variety of complex models.

Assessment and Grading Policy

- homework assignments (20%)
- 2 in-class presentations (40%)
- 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 (20% 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. 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 $\varepsilon > 0$" 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!

Two In-class Presentations (40% of Final Grade):
Each student will select two recent journal articles (with the approval of the instructor) to present in class.

Final Exam (40% of Final Grade):
The final exam for the class has two parts: (1) a take-home exam at a similar level to the homework assignments, (2) a half-hour oral exam.

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. Part of your grade will depend on participation in class discussions.

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 should be given to the instructor.

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

Course Structure

Review of 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 topics week by week, time permitting)

1. **Review**
2. **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.
3. **Weak convergence of stochastic processes**. Brownian motion, Brownian bridge. Functional delta method. Glivenko–Cantelli and Donsker Theorems.
4. **Bootstrap methods**
5. **M-estimators and empirical processes**
6. **Argmax continuous mapping theorem**
7. **Contiguity, quadratic mean differentiability, local asymptotic normality.**
8. **Martingale and counting process methods in survival analysis.** Kaplan–Meier and Nelson–Aalen estimators. Cox proportional hazards model.
9. **Empirical likelihood.** A nonparametric method based on a data-driven likelihood ratio function that does not require the user to specify a family of distributions for the data.
10. **Large p small n problems:** variable selection, regularization, non-standard asymptotics.
11. **Semiparametric inference and efficiency.** Score functions, estimating equations, tangent spaces, one-step estimators, double robustness.
12. **Post-selection inference.** Inference after variable or model selection can lead to inflated Type-I error rates or under-coverage of confidence intervals caused by "overfitting." Various methods of adjusting for such selection will be discussed.

Supplementary references

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*
Available as an e-book from Springer Link at
http://link.springer.com/book/10.1007/978-0-387-74978-5/page/1

Pollard D, *Asymptopia*. Book manuscript: http://www.stat.yale.edu/~pollard/

Owen A, *Empirical Likelihood*.
See also Owen's Fields Institute Seminar on Empirical Likelihood:
http://www.ms.uky.edu/~mai/sta709/Owen2005.pdf

Groeneboom P, 2013 Wald lectures, *Nonparametric Estimation under Shape Constraints*
http://dutiosb.twi.tudelft.nl/~pietg/Wald1.pdf

Martinussen T, Scheike TH, *Dynamic Regression Models for Survival Data*.
Available as an e-book from Springer Link at
http://link.springer.com/search?query=martinussen+scheike&facet-content-type=%22Book%22

Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models based on Counting Processes.* Available as an e-book from Springer Link at
http://link.springer.com/search?query=andersen+gill&facet-content-type=%22Book%22

Aalen OO, Borgan O & Gjessing HK. *Survival and Event History Analysis. A Process Point of View.*
Available as an e-book from Springer Link at
http://link.springer.com/book/10.1007/978-0-387-68560-1/

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.