Empirical Processes and Semiparametric Inference

Instructor: Professor Michael R. Kosorok, 3101 McGavran-Greenberg (McG).
Time and location: Mondays and Wednesday, 1:00–2:15PM, in room 1305 McG.
Office hours: Mondays 2:30-4:30PM.
Prerequisites: BIOS 761 or equivalent or consent of instructor.

Summary: Although theory of statistical inference has achieved a certain maturity in many fields of investigation, increasingly complex statistical models are emerging in clinical research, genetics, economics, atmospheric sciences, and in other areas. Over the past decade, developments in empirical process theory have proven to be powerful tools for working with these complex models. Semiparametric models, such as the Cox proportional hazards model or generalized additive models, consist of both a parametric and a nonparametric (infinite dimensional) component. These models are flexible and appealing because very few assumptions are made on the nonparametric part. However, this increased flexibility makes distribution theory quite challenging and modern empirical process techniques are usually required. This is currently an extremely active as well as demanding area of statistical research.

In addition to examining numerous examples, the course will develop in each student the technical skills to enable application of empirical process and semiparametric methods in statistics. Other areas to be covered include stochastic convergence in metric spaces, Brownian motion and Brownian bridges, Gaussian processes, Glivenko-Cantelli and Donsker theorems, entropy calculations, bootstrapped empirical processes; the functional delta method, Z-estimators, M-estimators, rates of convergence, semiparametric efficiency, semiparametric estimating equations, and nonparametric maximum likelihood estimation.

Required text: Kosorok (2008), *Introduction to Empirical Processes and Semiparametric Inference*, Springer-New York

(see `www.bios.unc.edu/~kosorok/`).

Supplemental texts:

V: van der Vaart (1998), *Asymptotic Statistics*, Cambridge University Press.

VW: van der Vaart and Wellner (1996), *Weak Convergence and Empirical Processes: With Applications to Statistics*, Springer.

Course requirements: The grade for the course will depend on regular homework assignments. 25 homework problems (in 11 problem sets) will be assigned throughout the semester to facilitate learning the material (worth 2 points per problem). There will also be three miniprojects that involve (1) finding a recent article on a related topic and describing what the researchers have done (10 points possible), (2) verifying some of the results (20 points possible), and (3) identifying interesting areas for future research (20 points possible). Miniproject (2) can be replaced by developing answer keys in latex for 5 additional homework problems that will be assigned separately for those interested in this option. The homework assignments and miniproject descriptions are given below. The final miniproject (MP3) must be turned in by 1:00PM on Monday, April 28. Grades will be determined on the basis of the number of points achieved (out of 100 possible): H (90–100%), P (80–90%), L (70–80%), and F (less than 70%). There will be no in-class exams.

Schedule of topics, chapters, and due dates. The assignments are due on Monday at the beginning of class for the given week, except for the weeks of January 13 and 20, in which case the assignments *will be due on Wednesday for those weeks. "PS" stands for problem set and "MP" stands for miniproject. The final assignment, miniproject 3 (MP3), is due Monday, April 28, at 1:00PM in my biostatistics department mail box.*

Week	Dates	Topic	Chapters	Due
1	Jan. 8	introduction and overview	1	
2	13, 15	"	2	PS1
3	[20], 22	"	3–4	PS2
4	27, 29	metric spaces, etc.	5–6	PS3
5	Feb. 3, 5	stochastic convergence	7	PS4
6	10, 12	empirical process methods	8	PS5
7	17, 19	entropy calculations	9	PS6
8	24, 26	bootstrapping empirical processes	10	MP1
9	Mar. 3, 5	the functional delta method	12	PS7
10	[10], [12]	(spring break: no class)		
11	(17), (19)	Z-estimators	13	
12	24, 26	"		PS8
13	31, Apr. 2	M-estimators	14–15	MP2
14	Apr. 7, (9)	Hilbert spaces, etc.	16–17	PS9
15	14, 16	semiparametric inference	18	PS10
16	21, 23	"	19–20	PS11
17	28	(assignment due at 1:00PM)		MP3

Miniprojects:

MP1: For this project, choose a recent statistical article (published since 2006) which utilizes empirical process methods and/or semiparametric inference. The major statistical or probability journals should be used. Please try to find a paper that involves a topic which seems to have significant future research potential, write a 2–3 page summary of the paper (you do not have to verify any math), paying particular attention to the practical issues being addressed. The paper(s) need to be approved on or before February 3, 2014. The total points possible are 10, and this paper is required of all students.

MP2: For this project, there are two options. For option 1, choose a recent statistical article (published since 2006) which utilizes empirical process methods and/or semiparametric inference. You may use the same paper you used for MP1. Please try to make sure that the paper contains significant mathematical-statistical content and involves a topic which seems to have significant future research potential. Identify 1–3 key steps in the proofs of the results which require empirical process and/or semiparametric techniques and then verify those steps (I want you to show me that you understand the steps involved). Write a 2–3 page summary of your verification of these steps along with any other insight into the workings of the technical aspects of the paper which you think are interesting. One question to ask is whether the author(s) could have shown the same results using simpler methods. You may want to get an early start on this paper and meet with me several times to make sure you are comfortable with the main technical aspects of the selected paper (I am quite happy to check your work and give clues if needed). The second option (option 2), is to prepare an answer key in latex for 5 homework problems selected by the instructor. Students interested in this homework option need to let the instructor know by February 3, 2014. The total points possible for MP2 are 20.

MP3: For this project, choose a recent statistical article (published since 2006) which utilizes empirical process methods and/or semiparametric inference. You may use the same paper(s) used for projects 1 and 2. Make sure that the chosen paper can provide some information about potential research projects. Identify 1–3 promising problems and/or research questions which could be of interest to the statistical community and which involve empirical process methods and/or semiparametric inference. Write a 2–3 page summary of your findings and include an evaluation of the potential impact if the proposed research were successful. The total points possible are 20, and this paper is required of all students. You may want to get an early start on this paper and meet with me several times to make sure you are headed in the right direction.

Problem Sets: All problems are taken from Kosorok (2008):

Assignment	Exercises
PS1	2.4.1, 3.5.1
PS2	4.6.1, 4.6.8
PS3	6.5.8, 6.5.13
PS4	7.5.5, 7.5.7
PS5	8.5.1, 8.5.7
PS6	9.6.6, 9.6.10
PS7	10.5.1, 12.3.2
PS8	13.4.2, 14.6.9, 14.6.10
PS9	15.6.2, 17.4.4, 17.4.5
PS10	18.5.3, 18.5.4
PS11	19.5.3, 19.5.7, 20.3.4

Reference texts: The following additional reference texts may be useful from time to time:

1. Andersen, Borgan, Gill, and Keiding (1993). *Stochastic Models Based on Counting Processes.*

2. Bickel, Klaassen, Ritov, and Wellner (1993). *Efficient and Adaptive Estimation for Semiparametric Models.* (abbreviated BKRW)

3. Ethier and Kurtz (1986). *Markov Processes: Characterization and Convergence.*

4. Fleming and Harrington (1991). *Counting Processes and Survival Analysis.*

5. van de Geer (1999). *Empirical Processes in M-Estimation.*

6. Pollard (1984). *Convergence of Stochastic Processes.*

7. Pollard (1990). *Empirical Processes: Theory and Applications.*

8. Shorack and Wellner (1986). *Empirical Processes with Applications to Statistics.*