BSTA 622 Statistical Inference II
Fall 2024

Content:
This course focuses on classical statistical inferential methods, including maximum likelihood estimation (MLE) and semiparametric estimation methods. We will study the asymptotic theory of MLE, unbiased estimating equations, and various likelihood estimation methods including composite, profile, and empirical likelihood estimation. This course will emphasize understanding the concepts, methods and theories by engaging with the most recent literature. Successful completion of this course will provide a solid foundation in probability-based statistical inference.

Intended Audience:
The course is designed for Biostatistics Ph.D. students in their 2nd year or beyond. Students are required to complete Probability I (BSTA 620) and Inference I (BSTA 621) before taking this course. Exceptions may be made with permission of the instructor.

Instructors:
Jing Huang, PhD
Jing14@pennmedicine.upenn.edu
Jinbo Chen, PhD
jinboche@pennmedicine.upenn.edu

TA:
Kelliher, Jimmy M
Jimmy.Kelliher@Pennmedicine.upenn.edu
Office Hours: TBD

Class Schedule:
Tuesday and Thursday 10:15-11:45 am in Blockley Hall 701.

Textbooks:
Recommended, not required, textbooks:
Theory of Point Estimation, by E.L. Lehmann and G. Casella, Springer
Elements of Large-Sample Theory, by E.L. Lehmann, Springer
Asymptotic Statistics, by A.W. van der Vaart, Cambridge
Theoretical Statistics, by D. Cox and D. Hinkley, Chapman and Hall

Grading:
The goal of this course is to introduce students to classic topics in statistical inference, familiarize them with seminal literature, inspire deep discussions to understand the theories, and equip them with foundational skills for theoretical research. Therefore, this course will emphasize understanding over examination. Grades will be determined by class interactions, discussions, and presentations.

Paper reading and presentation:
In addition to lectures, there will be four "paper reading and discussion" sections in this course. Each section will closely align with the topics covered in the class materials. Each section consists of two classes: one for paper reading and the other for paper presentation.

Paper Reading Class:
Only students and the TA will participate in the paper reading session. Students should arrive having already read the paper and formulated questions. They should also research related literature, either to aid their understanding of the paper or to broaden their grasp of the topic, prior to attending the paper reading class. This session will be an opportunity for them to discuss, pose questions, and engage in discussions with their peers.

Paper Presentation:
After the reading class, students will decide on the presentation format for the paper. During the presentation class, half the students will collaboratively present the paper. Every student is expected to participate actively in speaking. The presentation should be comprehensive, covering the paper's background, key contributions, theoretical justifications, implications, and potential extensions.

The grade will be based on performance in the paper presentation class and participation during the paper reading class based on the TA’s assessment.

Tentative Schedule

Date		Topics
Aug	27	Intro and Mathematics Primer
	29	Large Sample Theory
Sep	3	Asymptotic Properties of the MLE (Consistency)
	5	Asymptotic Properties of the MLE (Normality)
	10	Asymptotic Properties of the MLE (Efficiency)
	12	Paper reading: Self, S.G. and Liang, K.Y., (1987). Asymptotic properties of maximum likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association*, *82*(398), pp.605-610.
	17	Asymptotic Theory of Estimation
	19	Influence Functions
	24	Influence Functions
	26	Paper presentation Q&D for all the materials covered so far
Oct	1	Statistical Information
	3	Fall Break no class
	8	Statistical Information
	10	Paper reading: Varin, C., Reid, N., & Firth, D. (2011). An overview of composite likelihood methods. *Statistica Sinica*, 5-42.
	15	Generalized Linear models
	17	Paper presentation Q&D for all the materials covered so far
	22	Hypothesis testing: Likelihood ratio test, Score test, and Wald test (Reference: Wald, likelihood Ratio, and Lagrange multiplier tests in econometrics, Engle 1984)
	24	Hypothesis testing: Likelihood ratio test, Score test, and Wald test
	29	Profile likelihood estimation (Reference: Murphy and Van der Vaart. On profile likelihood. JASA 2000, 95:449-465)
	31	Instructor-led paper discussion: 1) Scott and Wild, Fitting regression models to case-control data by maximum likelihood. Biometrika 1997, 84:57-71. 2) Chatterjee N, Carroll R, Semiparametric maximum likelihood estimation exploiting gene-environment independence in case-control studies. *Biometrika* 2005, 92: 399–418. 3) Lawless et al., Semiparametric methods for response-selective and missing data problems in regression. JRSSB 1999, 61:413-438. 4) Chen and Little, proportional hazards regression with missing covariates. JASA 1999, 94:896-908.
Nov	5	Unbiased estimating functions/Estimation under model mis-specification (Reference: White, Maximum likelihood estimation of misspecified models. Econometrika 1982, 50:1-25.)

	7	Unbiased estimating equations
	12	Instructor-led paper discussion: Gronsbell J, Liu M, Tian L, Cai T. Efficient Evaluation of Prediction Rules in Semi-Supervised Settings under Stratified Sampling. JRSSB 2022, 84:1353-91.
	14	Paper reading: Minnier J, Tian L, and Cai T (2011). A perturbation method for inference on regularized regression estimates. JASA 106, 1371–1382.
	19	Paper presentation Q&D for all the materials covered so far
	21	Empirical Likelihood Estimation (Reference papers: Owen A. Empirical likelihood ratio confidence intervals for a single functional. Biometrika, 75, 237–249 Qin J and Lawless J. Empirical likelihood and general estimating equations. Annals of Statistics,1994, 22:300-325; Qin and Zhang, Marginal likelihood, conditional likelihood and empirical likelihood: connections and applications. Biometrika 2005, 92:252-270.
	26	Empirical likelihood estimation
	28	Thanksgiving no class
Dec	3	Paper reading: Williamson et al., A general framework for inference on algorithm-agnostic variable importance. JASA 2021, 118:1645-58. Williamson et al., Nonparametric variable importance assessment using machine learning techniques. Biometrics 2021, 77:9-22.
	5	Paper presentation Q&D for all the materials covered in the course