Stat 992: Course Logistics and Overview of Prerequisites

Hyunseung Kang

2025-01-21

Key Items from the Syllabus

- **Course website**: My homepage (Stat 992, Spring 2025)
- **Target audience**: Ph.D. students in statistics
- **Office hours**:
  i. Walk-ins whenever I'm available (1245B Medical Sciences)
  ii. By appointment (e-mail me)
- **Grading**:
  i. Summarize one paper; see syllabus

Goal of the Course

The main goal is to prepare students for research in causal inference.

1. Build intuition behind causal inference (e.g. confounding, counterfactuals, missing data)
2. Learn how to **identify** causal estimands:
3. Learn how to **estimate/infer** causal estimands:

Prerequisites

You need to know probability, math stats, linear models, and statistical computing at the graduate level (e.g., Casella and Berger (2002), Lehmann (1999), Rizzo (2019)).

Specific concepts are listed below

A. **Probability**: conditional independence/probability/expectation, law of total expectation, convergence of random variables (e.g., LLN, CLT), rates of convergence, continuous mapping theorem

B. **Math Stats**: likelihood-based estimation and inference (e.g., MLE, Cramer-Rao), nonparametric two-sample tests (e.g., t-test, permutation test)

C. **Linear Models**: linear models (projection-based), GLMs

D. **Computation**: bootstrap, cross validation, designing simulations to numerically evaluate estimators and tests (e.g., bias, Type I error rate)

My Go-To Reference Books

1. Serfling (1980), Lehmann (1999), and Lehmann (2006) (Appendix)
    i. For CLTs/LLNs under finite and super-population setups
    ii. For properties of U statistics (i.e. rank tests)
    iii. Rates of convergence are described intuitively in Lehmann (1999).
2. Newey and McFadden (1994) (Sections 2,3,6)
    i. M-estimation with estimated nuisance parameter
3. Wooldridge (2010) (Chapters 1-5),
    i. Deriving asymptotic properties of regression estimators
4. Van der Vaart (2000)
    i. For semiparametric efficiency theory[1]
    ii. For properties of M estimators and empirical process theory.

[1]There are now great references to this: Alejandro's book, Hines et al. (2022), Kennedy (2022), and Newey (1990)

Other Prerequisites

1. Cauchy-Schwartz inequality, and the triangle inequality
2. Taylor series approximation
3. Multivariable calculus and basic real analysis
    i. Open/closed/compact sets
    ii. Inf/sup/liminf/limsup, norms
    iii. Definition of limits, continuous funciton, and derivative
4. Linear algebra
    i. Linear span, column space, rank of a matrix, inverse, determinants
    ii. Orthogonal projections

References I

Casella, George, and Roger L Berger. 2002. *Statistical Inference*. Duxbury press.

Hines, Oliver, Oliver Dukes, Karla Diaz-Ordaz, and Stijn Vansteelandt. 2022. "Demystifying Statistical Learning Based on Efficient Influence Functions." *The American Statistician* 76 (3): 292–304.

Kennedy, Edward H. 2022. "Semiparametric Doubly Robust Targeted Double Machine Learning: A Review." *arXiv Preprint arXiv:2203.06469*.

Lehmann, Erich Leo. 1999. *Elements of Large-Sample Theory*. Springer.

———. 2006. *Nonparametrics: Statistical Methods Based on Ranks*. Springer.

References II

Newey, Whitney K. 1990. "Semiparametric Efficiency Bounds." *Journal of Applied Econometrics* 5 (2): 99–135.

Newey, Whitney K, and Daniel McFadden. 1994. "Large Sample Estimation and Hypothesis Testing." *Handbook of Econometrics* 4: 2111–2245.

Rizzo, Maria L. 2019. *Statistical Computing with r*. Chapman; Hall/CRC.

Serfling, Robert J. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons.

Van der Vaart, Aad W. 2000. *Asymptotic Statistics*. Vol. 3. Cambridge university press.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data*. MIT press.