

Stat 992: Course Logistics and Prerequisites

Hyunseung Kang

2024-04-03

Key Items from the Syllabus

- ▶ **Course website:** Canvas and my homepage
- ▶ **Target audience:** Ph.D. students in statistics
- ▶ **Office hours:**
 - i. Walk-ins whenever I'm available (1245B Medical Sciences)
 - ii. By appointment (E-mail: hyunseung@stat.wisc.edu)
- ▶ **Grading:**
 - i. One assignment. Submit the assignment by *May 3rd, 2024, 5:00pm Central*.
 - ii. See the course webpage for details.

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:
 - i. Under what conditions do we have
Causal Effect = g (observed data) for some function g ?
 - ii. Deals with population-level quantities (i.e. no randomness)
3. Learn how to **estimate/infer** causal estimands:
 - i. How should we estimate g , ideally with minimal assumptions?
 - ii. How should we test $H_0 : \text{Causal Effect} = 0$?
 - iii. Deals with randomness from sampling, experimental design, etc.
4. Learn how to conduct **numerical evaluations** for causal questions:
 - i. How do you simulate data for causal inference?
 - ii. What empirical metrics should you be looking for?
(e.g. covariate balance, overlap)

Probability Prerequisites (Non-Asymptotic)

You need to know probability at the level of an advanced statistics undergraduate student (e.g. Stat 309, Math/Stat 431, Ross (2010)).

1. Definition of conditional probability and conditional expectations
2. **Conditional independence**¹: If $X \perp Y | Z$, then for any functions f and g
 - i. $f(X) \perp g(Y) | Z$
 - ii. $\mathbb{E}[f(X)g(Y)|Z] = \mathbb{E}[f(X)|Z]\mathbb{E}[g(Y)|Z]$
 - iii. $\mathbb{E}[f(X)|Y, Z] = \mathbb{E}[f(X)|Z]$
3. **Law of total expectation**:
 - i. $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|Y]]$
 - ii. $\mathbb{E}[X|Y] = \mathbb{E}[\mathbb{E}[X|Y, Z]|Y]$

¹See Section 3.1 and Section 4 of Dawid (1979) for a concise list of implications arising from conditional independence.

Probability Prerequisites (Asymptotic)

- Limit theorems:** For $X_i \stackrel{\text{i.i.d.}}{\sim} F$ and F has finite mean and variance.
 - LLN: $n^{-1} \sum_{i=1}^n X_i \xrightarrow{P} \mathbb{E}[X_i]$
 - CLT: $n^{-1/2} \sum_{i=1}^n (X_i - \mathbb{E}[X_i]) \xrightarrow{D} N(0, \sigma^2)$
- Continuous mapping theorem:** For any continuous function $f(\cdot)$, if $X_n \xrightarrow{D \text{ or } P} X$, then $f(X_n) \xrightarrow{D \text{ or } P} f(X)$.
- Slutsky's theorem:** Let $Y_n \xrightarrow{P} c$ where c is a constant. If $X_n \xrightarrow{D \text{ or } P} X$, then $X_n Y_n \xrightarrow{D \text{ or } P} Xc$ and $X_n + Y_n \xrightarrow{D \text{ or } P} X + c$.

Math Stats/Stat Methods Prerequisites

You need to know math stats at the level of an advanced statistics undergraduate (e.g. Stat 310). Ideally, you should know math stat at the level of Casella and Berger (2002).

1. Generalized linear models (e.g. linear models, logistic regression)
2. Maximum likelihood estimators (e.g. efficiency, Fisher information, Cramer-Rao)
3. Hypothesis testing (e.g. Wald test, likelihood ratio test)
4. Parametric and nonparametric two-sample tests (e.g. two-sample t-test, Wilcoxon signed rank test, permutation test, etc.)

My go-to reference books: Serfling (1980), Newey and McFadden (1994) (Sections 2,3,6), Lehmann (1999), Wooldridge (2010) (Chapters 1-5), and Van der Vaart (2000)

Computational Prerequisites

1. You should know some R.
2. You should know how to simulate data and empirically evaluate
 - i. Properties of estimators (i.e. bias, variance)
 - ii. Properties of statistical tests (i.e. Type I error rate, power, coverage of confidence intervals)
3. You should know how to create reasonably informative plots or tables.

Other Prerequisites

1. Rates of convergence: $X_n = O_p(n^{-1/2})$ versus $X_n = o_p(n^{-1/2})$
2. Chebyshev's inequality, Cauchy-Schwartz inequality, and the triangle inequality
3. Taylor series approximation
4. Multivariable calculus and basic real analysis
 - i. Open/closed/compact sets
 - ii. Inf/sup/liminf/limsup, norms
 - iii. Definition of limits, continuous function, and derivative
5. Linear algebra
 - i. Linear span, column space, rank of a matrix, inverse, determinants
 - ii. Orthogonal projections

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. For M-estimation with estimated nuisance parameter
3. Wooldridge (2010) (Chapters 1-5),
 - i. For deriving asymptotics of regression estimators
4. Van der Vaart (2000)
 - i. For semiparametric efficiency theory²
 - ii. For properties of M estimators and empirical process theory.

²There are now great references to this: Alejandro's book, Hines et al. (2022), Kennedy (2022), and Newey (1990)

References

- Casella, George, and Roger L Berger. 2002. *Statistical Inference*. Duxbury press.
- Dawid, A. P. 1979. "Conditional Independence in Statistical Theory." *Journal of the Royal Statistical Society. Series B (Methodological)* 41 (1): 1–31.
- 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.
- Newey, Whitney K. 1990. "Semiparametric Efficiency Bounds." *Journal of Applied Econometrics* 5 (2): 99–135.
- Newey, Whitney K, and Daniel McFadden. 1994. "Large Sample