How to thrive in grad school
Session 3: How to read a paper

Y. Samuel Wang

Fall 2021

Introductions

What is a good book, TV show, or movie that you've watched recently?

Roadmap and expectations

We will cover

- Session 1: Basics
- Session 2: LaTeX
- Session 3: Computing
- **Session 4: How to learn from previous research**
- Session 5: Starting your own research
- Session 6: Communicating your research to others
- Session 7: Possible career paths

How to learn from previous research

Research is primarily disseminated in two main forms

- Papers
- Talks

Venues to find interesting/helpful papers

- Top 4 statistics journals: Annals of Statistics, Journal of the American Statistical Association (T&M), Journal of the Royal Statistical Society (B), Biometrika
- Theory focused journals: Bernoulli, Electronic Journal of Statistics
- Applications focused journals: JRSS C, JASA (A&C), Annals of Applied Statistics
- Machine Learning: Journal of Machine Learning Research, NeurIPS, ICML, UAI, AISTATS
- Others: Bayesian Analysis, Biometrics, Journal of Computational and Graphical Statistics, Journal of Multivariate Analysis, Scandinavian Journal of Statistics, Sankhya, Statistica Sinica, Statistical Science, Technometrics, etc.
- Review papers: Annual Reviews

How to read through a a paper

Papers should (in my opinion) be read in multiple passes

- First pass (5-15 minutes)
  - Spend 10-15 minutes reading at a high-level
  - Title, Abstract, Intro
  - What is the main problem the authors intend to solve?
  - Is this paper similar to other things you're familiar with? What does the paper accomplish that is different?
  - Do they show theoretical results? Empirical results?
  - If theory is involved, what is the main punchline of their theory?
  - In the simulations, which methods do the authors compare their procedure to?
- After the first pass, ask if it's worth taking another pass
  - Not relevant
  - Better to read other background first

How to read through a a paper

- Second pass (30-60 minutes)
  - What is the main innovation of the paper? (i.e., what's the secret sauce?)
  - Carefully look through any algorithms, figures, or assumptions
  - What are the parts of the paper would be helpful to understand further? (proofs, implementation details, etc)
  - What references are helpful to read next?
  - Is the paper internally consistent? do the arguments they make actually make sense?
  - Is the paper externally consistent? Do the results of the paper make sense in context of other papers?
- Third pass
  - Read through the details of proofs in the supplement
  - Read through parts of the paper you thought would be helpful but you didn't understand the second pass

Where to really find papers

- Google Scholar https://scholar.google.com/: Search engine for academic papers
  - Searching by keywords and skimming titles
  - Following citation trails
- Arxiv https://arxiv.org/: Pre-prints which are typically posted before the paper is accepted, but remains an open source way to access papers
  - Access papers not hidden behind paywells
  - See latest papers prior to publication
  - Sign up for digest for a daily email of the newest postings
- Cornell Passkey to access online journals from off-campus

Seminars to consider

- Go to the department seminars as often as you can (even if you don't think it will be immediately relevant)
- Initially, it may be really tough to follow what's going on in a talk. That's okay.
- Consider seminars from other departments: ORIE, CS, iSchool, Comp Bio, Sociology, Econ

Example

Find the paper: High-dimensional graphs and variable selection with the lasso

Homework

- Sign up for the ArXiv digest
  - Instructions: https://arxiv.org/help/subscribe
  - Categories: https://arxiv.org/category_taxonomy
- Get Cornell Passkey:
  https://www.library.cornell.edu/services/apps/passkey