

2018 年 11 月 Harvard, Department of Statistics, PhD 2 年 上原雅俊

今学期から 1 年目にあった週三回あった英語の授業がなくなりましたが、Teaching が入ってきてそれ以上の負担になっています。宿題の採点をしていると、アメリカの大学生は宿題をすごいきっちりやってくるなという気がします。一緒に宿題する友達がいなければ、いい成績を取るのがきつそうだなと感じがします。おそらく自分だったら GPA は散々になっていたでしょう。後は授業と研究について書きます。

(授業)

多変量解析

多変量解析は化石の学問で簡単だと思っていたのですが、かなり煩雑な計算があって、奥が深いなと感じました。教科書は Anderson の An introduction to multivariate analysis を使うのですが、謝辞に日本人の名前がたくさん書いてあって、そういう時代もあったのだと思いました。

GLM (Generalized linear model)

GLM の授業です。テストの日しか出ていないですが、必修なので取っています。GLM のために 1 学期も費やす必要があるのかなという感じがします。

Causal inference

Harvard で最も難しい causal inference の授業です。素晴らしくいい授業でした。Judia Pearl と Jamie Robins という因果推論界の二大巨頭が作ってきた Graphical model ベースの因果推論について学びます。巷にある教科書だと NPSEM と Do calculus ぐらいまでのことのしか書いていないですが、FRCISTG(有名な NPSEM とは違う causal model)、ID algorithm(causal effect を同定するための必要十分な ID algorithm)や Dynamic treatment regime における ID algorithm や SWIG(Single world intervention graph)など、新しいことをたくさん学べました。研究で使えたらいいなと思います。

High dimensional Statistics

これは授業というより友達とやっているセミナーです。Roman Vershynin の High dimensional probability と Martin J. Wainwright の High dimensional statistics という二つの本を問題と一緒に解きながら輪読しています。どっちとも素晴らしい本です。大分、集中不等式や経験過程に基づいた現代的な高次元統計(学習理論)に慣れてきました。

Causal inference の勉強会

Harvard の先生が主催している二つの Causal inference の勉強会に時々出ています。大体の話についていけるようになってきたので、来年くらいから研究できたらと思います。

2018年11月 Harvard, Department of Statistics, PhD 2年 上原雅俊

(研究)

今年になって論文を2本 submit しました。一つは学部のころやっていた研究で、二つ目は Iowa State University で夏にやっていた欠測データの研究です。

<https://arxiv.org/abs/1808.07983>

<https://arxiv.org/pdf/1810.12519.pdf>

Iowa State University では Kim 先生という方と共同研究しました。うまくマッチしてすぐ論文が書けました。欠測データや Causal inference という分野では大抵 Missing at Random という仮定を置きます。今回の研究は Nonresponse IV(操作変数)とセミパラメトリックな Response model(Propensity score model)を仮定して Not missing at random 下における推定手法を考えたという研究です。今はこの論文と他の関連した論文の Semiparametric efficient estimators について時間のある時に考えています。ただこれ以上、セミパラ系の欠測やってもやり尽くされているので、もっと新しいテクを入れなければという気持ちを強くしました。

他に結構前からやっている計算統計とセミパラメトリック統計の融合的な研究をしています。とりあえず、二つの論文を完成させるべく、頑張っています。一つの研究は Unnormalized Model の efficient estimators の研究で、日本に5月に帰った時、東工大の先生のところで発表した関係で、興味を持ってもらえて共同研究しています。研究内容を説明すると長くなるので出来ませんが、Keyword は Bregman divergence, Gamma divergence, Nonparametric estimator, Unnormalized models です。共同研究はやっぱり楽しいなと思います。もう一つは Bayesian Monte Carlo integration の話です。Monte Carlo integration は一般的に Frequentist 的なのですが、それをベイズ的にやるという話です。漸近論的にとても面白い結果が言えて、解析的な形式に頼らず Uncertainty quantification ができることが理論的に言えます。また詳しく書けたらと思います。

最近、始めたのがネットワークモデルの理論研究です。正直、特別にネットワークという分野に研究に興味あるわけではないのですが、高次元統計や学習理論のテクを手に入れるモチベーションで始めました。(あともう簡単なモデルはやり尽くされているのでネットワークぐらいしかない?) プリンストンから来た若い助教の人とやっているのですが、時間取ってくれるので、効率よくできていると思います。ただこの領域、理論の強い人がうじゃうじゃいるので、関連論文も難しいです。

今までは研究成果というよりは、欲しい武器がいかにか得られるかというスタンスで研究してきたのですが、大分、増えてきたので、いよいよ来年からは PhD を通してやるテーマ(多分 causal inference と reinforcement learning)に焦点を当てて研究したいなと思います。