

船井奨学金財団報告書

Harvard 大学 PhD1 年の上原雅俊です。出願したときは、統計に正直、興味を失いつつあったのですが、最近は統計学が面白いなと感じるようになって良い精神状態を保っています。というのも周辺に応用のことをやりつつメソドロジーの研究もやっている人が結構いると分かったからだと思います。統計学では解きたい科学的なモチベーションが先あって、データ解析を行ったり、手法を考える分野を一般的に応用統計といいます。論文を出す雑誌としては応用統計の雑誌であったり、それぞれの分野で良い雑誌だったりします。一方、個別のデータ解析というより、それなりに一般的な状況で使える手法を考え、その理論保証をするのがいわゆる統計学です。(勿論、境界は曖昧なものですが) 統計学者でコンセンサスが取れている良いジャーナル、4 大誌というのがあって、アカデミアを目指す場合、そこに論文がいくつあるかというのがキャリア形成に重要になってきます。応用的なことではっきりある特定の分野に貢献し、先ほど言及した 4 大誌にボンボン載せている教授が Harvard の統計学科や生物統計学科には結構います。そういうことができたらいいなというロールモデルをすぐ間近に見ることができて刺激を受けます。残りは授業と研究について主に書こうと思います。

授業

Statistical inference 2.... 必修の統計の授業です。内容は漸近論、モデル選択、MCMC、高次元統計とか統計研究する人だったら知っているべきことを薄く広くやる感じでした。

Bayesian Data analysis 必修のベイズ統計の授業です。先生はベイズ統計で有名な人なのですが授業は微妙な感じでした。内容は階層モデル、MCMC、ノンパラベイズとかでした。

Probability 2.... 必修の確率論の授業です。MIT の数学科を退官した教授が教官でした。夏学期にも確率論の授業はあったのですが、今回の授業はもっと厳密なものでした。

High dimensional Statistics and probability MIT で行われている高次元統計と確率論の授業を聴講しました。先生は学習理論で有名な人で、中身はかなり理論的なものにもかかわらず人気でした。

TA 養成のための授業 1 ...TA 養成のための統計学科の授業でした。

TA 養成のための授業 2 TA 養成のための GSAS(大学院の組織)の授業です。毎週何か自分の分野の Term を知らない人向けにプレゼンする授業があります。この授業まで

PhD で自分より英語できない人を見たことがなかったのですが、今学期は何人かそういう人を見かけて少し安心しました。

研究

今は **Unnormalized Model** という規格化定数が計算できないモデルの推定手法の研究をアドバイザーとしています。最近から **IMS**(国際数理統計学会)の会長をやっている忙しいので、自分からどんどんプッシュしていかなければならないのですが、会った時は色々アドバイスをくれるので効率的に研究できています。**Unnormalized Models** の応用例としては **Bohlmann machine**、切断回帰モデル、独立成分分析などがあります。このモデルには色々な推定手法が考えられるのですが、いかに漸近分散(エラー)を下げるかということを目的にして、二つの観点から新しい手法を生み出すようにしています。一つの研究は **Unified Estimation Methods for Unnormalized Models Using Auxiliary Distributions** というタイトルで今年の 8 月の **JSM**(アメリカの統計学会)で発表する予定です。もう一つの手法はノンパラメトリックモデル(無限次元のモデル)をプラグインする手法でセミパラメトリックモデルの漸近解析がメインです。因果推論を学ぶために去年、セミパラメトリックモデルを先学期かなり勉強したので、勉強したことが使えるのは楽しいです。**Unnormalized models** の研究と関連してモンテカルロ積分にいかにか事前情報(不等式の情報や他のソフトウェアから得た情報)をベイズ的な視点と頻度論的な視点で埋め込むかという研究もしています。詳しくはまた今度、触れられたらいいなと思います。

最近では欠測データの研究も始めました。博士課程では上の研究よりはこの二つを中心にやっていこうと思っています。因果推論はどのように因果効果を測定するか、欠測データは欠測データをいかに扱うかという話です。一見両者は関係なさそうですが、かなり重なっている部分が多いです。学部生の頃はどちらも興味なかったのですが、実際どのようなところに統計が生きているかというのを考えているうちにこの二つに興味が湧いてきました。例えば疫学や社会科学で因果関係を導くというのは、中心的なモチベーションです。まだデータが欠測しているというとネガティブなイメージに一見聞こえますが、欠測データは避けられないし、あらゆることが欠測データの枠組みで捉えることができ、因果推論、潜在変数モデル、標本調査論といった分野の研究は欠測データの研究と密接に関係しています。

欠測データの研究は **Harvard** の生物統計学科の先生と研究したかったのですが他の大学に移勤するということで断られたので、**Iowa State University** にいる教授と研究す

ることにしました。Iowa State University は標本調査論のメッカで、標本調査論と欠測データで有名な先生と研究を6月から開始しました。やっている内容は Semiparametric propensity score model を仮定して、Not Missing at Random 下の欠測データを扱う研究です。詳しくはまた今度、触れられたらいいなと思います。研究結果は JSM で発表される予定です。一方で、因果推論の研究はまだアドバイザーを探しています。一度、因果推論と強化学習をやっているアドバイザーの元で去年始めたのですが自分のやりたいことと違っているということですのですぐ辞めた経緯があるので、誰がいいか今、探っています。来学期の目標は今やっていることを仕上げつつ、また新しい研究を始めることです。持っている武器ももっと増やせるようにしていきたいなと思います。