

先学期のことと、夏のことについて書こうと思います。先学期は、授業は受けずに、Teaching だけしていました。

- Teaching

Introduction to stochastic process という確率過程の授業を教えました。やったことは週2回の office hour、週2回の授業、宿題・中間・期末の添削です。自分が大学生の時に、こんなに色々やる TA がいたらよかったと思います。

- 授業(聴講)

授業は取っていないのですが、MIT で High dimensional statistics の授業を聴講しました。Martin Wainwright の High dimensional statistics という最近出版された素晴らしい本に沿った授業です。授業する人が学習理論で有名な人で、とても分かりやすかったです。もう一つ、stochastic process の授業を聴講しました。将来的に、確率過程の統計もやれたらという憧れがあるのですが、まだまだかなり修行が必要そうです。

- ゼミ

友達と Bandit problem についてのゼミを Regret Analysis of Stochastic and Nonstochastic Multi-Armed Bandit Problems という本でしていました。Bandit problem は、簡単に言えば損失 (Regret) を最小化するように実験のデザインを考えることが目標です。Regret をいかに不等式で抑えるかが肝なのですが、未だによく気持ちが分かっていません。今年、Bandit の 500 ページを超える今まで一番詳しい本が出るので、それも読めたらなと思います。

- 夏の研究

5-8 月の中旬まで New York の Cornell tech で研究していました。Cornell tech は Cornell University の OR・CS department の Faculty 達がいる分校を指します。New York の Manhattan 島の隣の島にあって、便利なところにあります。受け入れてくれたのは Nathan という機械学習・因果推論をやっている先生です。その方は元々、最適化が PhD の時の専門なのですが、統計とか学習理論もかなりわかっていて、統計や機械学習のいい journal・conference にもかなり通しています。いわゆる幅の広さで新たな視点を得て勝負する系の人で、自分も見習っていけたらと思います。

・研究

前回の報告書から 4 本、論文書いたのだからそれらについて簡単に書きます。

### **Double Reinforcement Learning for Efficient Off-Policy Evaluation in Markov Decision Processes**

<https://arxiv.org/abs/1908.08526>

Off policy evaluation(OPE)についての Cornell の先生との論文です。OPE とは過去のデータから方策の効果を推定する問題のことを指します。例えば薬や、マーケティング効果などの測定が応用例としてあります。これは主に Non Markov Decision Process と (time-variant) Markov Decision process における OPE の推定下限を導出し、それを達成する推定量を達成したという論文です。特に MDP において下限を達成する推定量は意外に存在しなかったのだから、そこに新規性があります。これが書き終わるまで集中するために色々なことをシャットアウトしてしていました。(報告書の提出が大幅に遅れた口実として...) 今までで一番いい出来だと思います。もう一つ Time-variant MDP の論文も書いていて、それも早く終わらせた方がいいなと思います。

### **Intrinsically Efficient, Stable, and Bounded Off-Policy Evaluation for Reinforcement Learning**

<https://arxiv.org/abs/1906.03735>

これも同じ Cornell の先生との OPE の論文です。この論文で3つのよく知られている推定量より漸近的に優越になるという Intrinsic efficient な推定量を導出しました。ただ因果推論で知られている話を、強化学習に持ってきたという感じなので、因果推論の理論に詳しい人から見たらそれはそうという感じだと思います

### **Unified estimation framework for unnormalized models with statistical efficiency**

<https://arxiv.org/pdf/1901.07710.pdf>

Unnormalized Model においては MLE という統計の Golden standard とも言える手法が使えないのですが、その性質を保持できる推定量を divergence の観点から幾つか提案したという論文です。東工大、はこだて未来大、東大の先生との論文です。論文の分類的にはセミパラメトリック統計、情報幾何、計算統計になるのかなと思います。結果自身はそこそこ面白いと信じているのですが、絶妙なバランスの予備知識が要求されるので、非常に読みにくいです。

## Imputation estimators for unnormalized models with missing data

<https://arxiv.org/pdf/1903.03630.pdf>

欠測の論文を書きたかったので、書いた論文です。東大と Iowa State University の先生との共著論文です。EM algorithm、Fractional imputation と Multiple imputation という欠測データの解析で知られている手法を Unnormalized model の推定量と組み合わせ、かつ理論的な性質を調べたという論文です。欠測データのアルゴリズムの理論を勉強したので、それをただ使ってみたいというモチベーションで書きました。

最後に

来学期からイリノイにある UIUC の CS department で 1 年間、強化学習の研究をおこなうに行きます。今のところ強化学習の Exploration の理論をやる予定です。ただこの分野は理論系ながら、大学にいた人が企業に移って研究していて、彼らがとても強いです。最初、入りにくそうな領域なのですが、一旦軌道に乗ったら楽になることを信じて頑張ろうと思います。最後になりますが、船井情報科学振興財団の皆様方、2 年間のご支援ありがとうございました。金銭的な余裕が精神的な余裕にもつながり、研究に打ち込むことができました。これからもさらに精進していけたらと思います。