

留学報告書 Part 1

こんにちは！現在UCバークレーの統計学科に通っていますPhD1年の石原みやびです。前回の報告書では大学院に入るまでの経緯について書きましたので、今回は前期の学生生活、授業についてお伝えしたいと思います。

UCバークレー統計学科の初日



バークレーの時計台 Sather Tower. 右手の建物はDoe Libraryです。手前の芝生では学生がフリスビーやクディッチをしたり、寝転がって本を読んだりします。

2017年8月21日午前8時半。UCバークレー初日、統計学科のオリエンテーションが行われるラウンジに着くと最初に目に入ったのはテーブルの上にあるクロワッサン、フルーツ、コーヒー。朝食に手がつけられずソワソワしていると見慣れた顔が集まってきました。3月の大学訪問で会った人が多いので自己紹介はほどほどに、話題は家探しの苦労話で盛りあがります。

PhDの新入生は10人。学部・修士課程を卒業してPhDに進学した人、数年間企業で働いていた人や研究のため1年間のギャップイヤーを経てきた人がいます。学部での専攻は、数学・応用数理・統計学のいずれかですが、中にはダブルメジャーやマイナーでコンピューターサイエンス、生理学、英文学を学んでいた学生もいます。留学生は3人。出身はロシア、フランス、日本です。

ほどなくして職員の La Shanaと学科長の Steve 教授から、バークレー統計学科の成り立ち、PhDの卒業要件について話がありました。実はこの日、午前10時に北アメリカ大陸を横断する皆既日食をみる予定でしたがあいにくの雲り空で見ることができなかったため、教授に今学期履修する授業を確認してもらい、その後みんなで教員行きつけのメキシカン料理店でランチをしました。夕食はいらないと思えるくらい大きなブリトーでお腹がパンパンになります。

大学に戻ってからは図書サービス、パソコン施設、統計クラブについての説明を休憩なしで一気に受けます。全ての新入生に義務付けられているセクシャルハラスメント・性暴力防止のワークショップでは、Tea Consentという3分のアニメーション動画を題材に、お互いの『同意』は尊重した人間関係を築くために大事であると学びました。夕方は教員、職員、学生の立食パーティー。生春巻きと焼き鳥のケータリングを見たら、いらないと思っていた夕食ですがやっぱり食べてしまいます。

家に着いたのは午後7時。1日目が無事に終わって、緊張感から解放されてぐっすりでした。



Sather Tower の展望台からはキャンパスとその先のサンフランシスコのゴールデンゲートブリッジを見ることができます。

履修計画

PhDの1年目は、統計学科の基礎科目である確率論、理論統計、応用統計の全7クラスから年間最低3クラスを履修しなければなりません。NYUの修士課程に在籍していた時は年間8クラス履修していたことを考えると3クラスは少ないように感じましたが、上級生からは前期2クラスで十分と言われたこともあり、理論統計と応用統計の2クラスを履修することにしました。実際に授業が始まってみると、1クラス週2日3時間の授業とその課題に20時間程度、演習が2時間、データサイエンスのゼミ DS421 で論文を読んでディスカッションするのに1時間、統計学者のゲスト講演に参加する Neyman Seminar が1時間とハードワークでしたから、2クラス履修は納得のアドバイスでした。

コースワーク

履修した理論統計と応用統計について少し詳しく紹介します。



Evans Hallにある Mathematics Statistics Library. 9万冊以上の蔵書があり、大学院生は最大200冊借りることができます。

理論統計では、Statistical decision theory, Point estimation, Hypothesis testing, Large-sample theory, Multiple testing (統計的決定論、指数型分布族、点推定、仮説検定、大標本論、多重検定)などを扱います。週3時間の講義に対し、証明問題を中心とした5つの大問が課題として出されます。すんなりわからない問題は(ほとんどですが)、TAのオフィスアワーを利用してヒントをもらったり、図書館で関連図書を読んだり、同期の何人かで集まりラウンジにある大きなホワイトボードを使って議論しながら解いていきます。また、情報交換サイト Piazza を活用する学生も多いので、講義や課題に関するディスカッションはネット上で頻繁に行いました。

例えば、仮説検定について学んでいる週は以下のような問いが出されましたが、この問題は特に難しく、他の学生と考えても明快な解答に行きつくことができませんでした。

This question concerns hypothesis testing in the Cauchy location family:

$$p_{\theta}(x) = \frac{1}{\pi(1 + (x - \theta)^2)}.$$

Prove that there exists no UMP test for testing $H_0 : \theta = \theta_0$ vs $H_1 : \theta > \theta_0$.

結局、解答の方針のみを提出し、翌週配られる解答例を読んで学ぶしかありませんでした。ただ、難しい問題ほど同期生との議論が多くなりますし、その分解答を読んだ時の“Aha!”も大きかったです。以前、バークレー統計学科の卒業生が言っていた“Part of assigning homework is to make everyone struggle”を思い出し、その言葉は的を得ていると感じました。

一方、応用統計は実際の社会・自然現象のデータを分析する授業です。授業の内容は、exploratory data analysis, visualization, linear regression, dimension reduction techniques, classification and clustering (探索的データ分析、ビジュアライゼーション、線形回帰、次元削減、分類とクラスター分析) です。これらはすでに修士課程で学んだことですが、担任のBin教授は分析手法のさらに深い理解を要求します。「このデータにこの分析方法を使うのは本当に妥当なのか」、「手法Aの代わりに手法Bを使うメリット・デメリットは何か」など、分析過程の各ステップをクリティカルに考える質問を投げかけます。

2週間に1度出されるデータ分析のプロジェクトではレポートを提出後、名前を伏せて学生同士でフィードバック (peer review) をします。オープンエンド型のプロジェクトは、正しい分析方法や答えがあるとは限らないので、他の学生の分析法や解釈の仕方から学べる peer review はとても貴重です。データ分析は小論文を書く作業と似ていると言われますが、同じデータを与えられてもどのようにビジュアル化し、分析するかは十人十色。分析結果をわかりやすい言葉で説明している同期生のレポートには刺激を受けます。

データの扱い

2017年日本では、大学教授による論文データの捏造疑惑が発覚するなど、アカデミアにおける不正問題が浮き彫りになりました。このニュースを受け、応用統計の授業ではデータの捏造問題について取り上げ、「どこまでがデータクリーニングとして許され、どこからがデータの捏造となってしまうのか」が話題になりました。

データクリーニングとは分析しやすい形にデータを整備することです。一方、捏造は特定の結果を得るために意図的にデータに手を加えることを指します。一見明らかに違う2つですが、外れ値を除外しすぎたりするなど、クリーニングのやり方を間違えると分析結果も本来のものとはずれてしまい、結果として周りからは捏造をしたと受け取られてしまうこともあります。今回問題となった画像操作の捏造疑惑に関しては「結論を操作しようとする意図ではなく、画像をわかりやすくするためのデータ処理のようなので、捏造とは思わない。ギリギリのラインであることは確かだけど」という意見もありました。

私が参加しているデータサイエンスDS421のセミナーでSol教授は「論文を世に発表したら、いつ誰が君が行ったデータの分析を掘り返すかわからない。その意味でも、分析をするときはとにかく一つ一つのステップを念入りに考えて行うべき」と学生に念を押します。しかし、と教授は続けます。「どんなにしっかり分析をしていたとしても、いざ分析内容について質問を受けるときはヒヤッとすると、万人が納得する分析を行う難しさにも言及します。

分析はマニュアル通りに行えるものではなく、多くの難しい決断を要するものだということを再認識しました。一方で、データ分析における曖昧さを軽減することも大事です。Bin教授が他の統計学者と共同で出している [Ten Simple Rules for Effective Statistical Practice](#) には、健全な統計解析を行うための心構え・やるべきことがまとめられていて参考になります。

ソーシャルイベント

授業以外に、人と人との交流を大切に作る“場”が多くあります。ともしれば、研究室に閉じこもりがちになる学生たちを食べ物で研究室の外に連れ出そうとスタッフが企画しているcookie hour は毎週火曜日にあります。ラウンジに行くと、大皿にチョコレートチップクッキーや大好きなブラウニーが可愛らしく整列していて、小腹を空かせた学生と教員が授業や研究について気軽に話すことができるうれしい時間です。

他にも統計クラブが企画して、ラウンジでカードゲームやテーブル・フットボール大会をすることもあります。また食事会が毎週金曜日の夜に大学近くのバーで行われ、学生同士が趣味や特技や意味のないことについて話します。ある時一人が、豚と馬のどっちになりたいかと質問した時には大いに盛り上がりました。彼の考えでは、「豚は一生食べ物を与えられひもじい思いをすることはないけれど丸々太ったら食用のため連れて行かれる。一方、馬は一生働かされ酷使されるけれど、食用のために殺されてしまうことはないでしょ」で、「さあ君ならどっちを選ぶ?」「……!?!」

このようなソーシャルイベントのメリットは、serendipity。思わぬ出会いと展開が研究に良い影響を及ぼすということでしょう。お互いを知ることによって、話しやすい雰囲気ができて、クラスメートというより同僚としての意識が育まれてきたように思います。

2017年を振り返って

大学構内で起きた非日常的な出来事が3つありました。2つは抗議活動で1つは大規模な山火事です（ただ、どの日も授業は普段通りに行われていました）。それぞれ簡単に説明します。

9月にはキャンパス内のホールで、保守派の政治コメンテーターであるベン・シャピロ氏のフリースピーチが開催されました。予想されるプロテストに対し、スピーチ開始の数時間前から大学周辺が厳重警備されました。全学生にはプロテスト区域を避けて移動するための安全な帰宅ルートがメールで送られてきたほどです。

11月には、大学院生の学費免除特典に課税するという政府の法案に対するデモでキャンパス内が人でごったがえしていました。大学生にとって直接関わる重大な問題でしたが、12月にこの法案は取り下げられました。

カリフォルニア州は長年干ばつに悩まされており、10月にはワイン産地として有名なソノマ郡やナパ郡などでカリフォルニア州過去最悪の山火事が起きました。この地域一帯で発生した煙は、90kmほど離れたバークレーキャンパスにも押し寄せ、太陽が煙で隠れる日が3日ほど続きました。クラス



毎週火曜日、午後3時に出されるクッキーとブラウニー。取りすぎに注意です。

メートの中には親戚の家が全焼し、今は別の親戚の家に避難している人もおり、長期にわたる被害であることを痛感しました。

統計学では稀な事象の生起確率を推測する研究分野（rare event modeling）がありますが、複雑な自然現象を予測するのは困難です。災害が起こるメカニズムを解明し、災害が起こってしまった場合、被害を最小限に抑えるためにどう対応すれば良いかを探っていきたいと思いました。

これから

2017年度前期はコースワーク中心で課題やプロジェクトに時間を多く使いましたが、後期は研究に関して教授と話す時間を増やし、タイムマネジメントをして活動の場を広げていきたいと思っています。また、人との出会いを大切に、新しい知識を吸収しながら成長していきたいと考えています。この報告書を読んだ方でバークレーあるいはベイエリアを訪れる機会がありましたらぜひ [miyabishihara\(at\)gmail.com](mailto:miyabishihara(at)gmail.com)までご連絡をください。

ハウスメイトの飼い猫

