Course title: Adaptive Experimentation: Bringing Real-World Challenges to Multi-Armed Bandits

Speaker: Chao Qin (Stanford Graduate School of Business)

Website: https://cqin211.github.io/

Email: chaoqin@stanford.edu

Abstract: Classical randomized controlled trials (RCTs) have long been a gold-standard method for causal inference and data-driven decision making in many fields. In recent years, adaptive experimentation has emerged as a promising alternative, with the potential to deliver greater efficiency and cost-effectiveness than conventional RCTs. While adaptive experimentation has been widely studied across communities (e.g., multi-armed bandits (MABs), reinforcement learning, Bayesian optimization, active learning, and online learning), its real-world deployment remains difficult due to a range of practical challenges.

This mini-course presents recent advances in adaptive experimentation, taking a step toward bridging the gap between academic research and practice application by incorporating real-world challenges into the MABs framework. These challenges include tradeoffs between competing objectives, hard experimentation deadlines, nonstationary environments, delayed feedback (e.g., batch settings), and exploration tasks that go beyond learning the best treatment arm.

Goals, importance, and timeliness: The mini-course aims to introduce practice-aligned models, algorithm design principles, and theoretical foundations likely to remain influential in the academic literature for years to come. Additionally, it seeks to provide practitioners with practical algorithms and actionable insights.

Meeting times and locations:

- Friday, May 2 | 3:15 PM – 5:15 PM CDT 2110
- Tuesday, May 6 | 3:15 PM – 5:15 PM CDT 5101
- Wednesday, May 7 | 10:00 AM – 12:00 PM CDT L070

All rooms are located within the Kellogg School of Management at 2211 N Campus Dr, Evanston, IL 60208.

Outline:

Lecture 1: Friday, May 2 | 3:15 PM – 5:15 PM CDT

- Introduction to RCTs and MABs
    - A core tension in adaptive information gathering: robustness vs. efficiency
    - Applications of bandit algorithms
- Tradeoffs between competing objectives (e.g., regret minimization vs. pure exploration)
    - A generalized model of adaptive experiments
    - Unification of canonical theory
    - Insights into the nature of universally optimal policies
    - Small adjustments to popular bandit algorithms achieve universal optimalityœ
    - Length-regret example: Pareto frontier of experiment length vs. total regret
    - *Readings: Qin and Russo [2024], Lai and Robbins [1985], Chernoff [1959], Garivier and Kaufmann [2016]*

Lecture 2: Tuesday, May 6 | 3:15 PM – 5:15 PM CDT

- Focus on best-arm identification (BAI)
    - A special example from Lecture 1
- BAI with adaptive soft stopping
    - Soft stopping, also known as fixed-confidence setting
    - Equivalent formulations of fixed-confidence setting in the asymptotic regime
    - Universal optimality is achievable
    - Other notions of optimality: posterior convergence, large-deviation rate
    - *Readings: Qin and Russo [2024], Glynn and Juneja [2004], Russo [2020], Kasy and Sautmann [2021]*
- BAI with hard stopping: No universal optimality
    - Hard stopping, also known as a fixed-budget setting
    - A ‘dual’ problem to fixed-confidence setting
    - No universal optimality for this setting
    - *Readings: Qin [2022], Bui, Johari, and Mannor [2011], Ariu, Kato, Komiyama, McAlinn, and Qin [2021], Degenne [2023]*
- BAI with hard stopping: Pursuit of universal dominating policies over RCTs
    - RCTs, also known as uniform allocation, A/B/n testing
    - Uniform allocation is inadmissible
    - Batched arm elimination (BAE): A class of RCT-dominating policies
    - Minimax BAE: a robust optimal one within the class
    - *Readings: Qin [2022], Imbens, Qin, and Wager [2025]*

Lecture 3: Wednesday, May 7 | 10:00 AM – 12:00 PM CDT

- Nonstationary environments, delayed feedback (e.g., batch settings), distribution shifts
    - A practical model that captures these challenges
    - Other popular bandit algorithms fail entirely, even in simple examples
    - A simple modification of Thompson sampling (TS) suffices: deconfound Thompson sampling (DTS)
    - DTS is as robust as RCT in complex environments, yet significantly more efficient (and optimal) in benign environments
    - *Readings:* *Qin and Russo [2023]*, *Qin and Russo [2022]*
- Exploration tasks that go beyond learning a best treatmenElectronic Companion to Adaptivity and Confounding in Multi-Armed Bandit Experimentst arm
    - A unified approach to a broader class of problems, extending top-two algorithm design principle
    - Illustration of how to adapt TS
    - Specialized for best-arm identification: resolving a notable open problem
    - *Readings:* *Qin and You [2024]*
- Other practical concerns – time permitting