Empirical Models in Marketing, Part B

Professor Ryan Dew
Office: JMHH 755
ryandew@wharton.upenn.edu

Overview

This doctoral seminar introduces students to advanced methodologies in empirical models for marketing. Building on Empirical Models Part A, which covers the foundations of empirical modeling, Part B is hands-on and forward-looking, covering models and methods that are emerging, with a goal of giving students a cutting-edge toolkit to build their own empirical models. In particular, we will approach model building, estimation, and inference from a probabilistic perspective, drawing on fundamental methodologies from Bayesian statistics and probabilistic machine learning, to study consumers and markets. We will focus on two broad topics: (1) modeling choice, including flexible and nonparametric specifications of heterogeneity and dynamics, scalable models for large assortments, and adaptive and reinforcement learning specifications; and (2) modeling unstructured data (e.g., text, images), including topic modeling, representation learning, and generative modeling.

Goals

Given the course is only six weeks long, coverage of any one topic will be somewhat brief. The goals of the course are:

1. to expose students to the exciting methodological developments in empirical modeling taking place at the intersection of econometrics, Bayesian statistics, and probabilistic machine learning, so that these topics can be pursued in greater depth in their own research;
2. to build a general purpose toolkit for probabilistic modeling and programming that can be used to implement a wide array of models; and
3. to connect these ideas to research in marketing, to hopefully spark new research ideas.

Format

Class sessions are a combination of lectures and student-led discussions of papers. Each session is divided into three roughly 1-hour long "mini-lectures," with breaks in between, and where the first hour is typically devoted to students presenting papers related to the prior week's topic. There will also be weekly assignments to reinforce basic concepts, and give students the opportunity to implement the ideas we discuss in class. These weekly assignments are always due prior to the start of the subsequent session. The course will use Canvas for organization.

Prerequisites

- It is recommended that students take Part A of the class before taking Part B.
- Knowledge of the basics of linear algebra, probability (e.g., chain rule of probability, basic probability distributions, Bayes rule), statistics (e.g., maximum likelihood) and choice modeling (e.g., multinomial logit and probit models) will be assumed.
- Assignments will involve coding; examples will be given in Python, but students may elect to use another language if they wish. (However, using Python is strongly recommended, and many packages we will use are Python-specific.)

Grading

- 50% weekly assignments
- 50% in-class engagement
    - On-going participation during in-class discussions
    - Leading a paper discussion

Getting Help

This class is a fast-paced, intense learning experience, aimed at maximize your learning during our short time together. I know that things will move quickly, and this always creates a risk of students falling behind. If you need help, whether it's with understanding a concept or a paper, or with one of the assignments, or anything else, please reach out — don't be shy! You can stop by my office any time you see the door open. I'll also post a link on Canvas where you can book a time to meet one-on-one with me, in-person or virtually. My goal is learning, not suffering!

Tentative Schedule

Note: This plan may be overly ambitious, and thus, potentially subject to change.

Session 1: Fundamentals of Model-building and Inference

- 1.1. Course intro + Basics of model building
- 1.2. Bayesian and frequentist approaches
- 1.3. Inference algorithms and probabilistic programming

Session 2: Dynamics and Heterogeneity I

- 2.1. State-space models and HMMs

- 2.2. Bayesian nonparametrics and the Dirichlet process
- 2.3. Mixtures of {Normals, DPs, HMMs}

Session 3: Dynamics and Heterogeneity II

- 3.1. Student-led discussion(s): DPs
- 3.2. The Gaussian process and dynamic heterogeneity
- 3.3. Factorization methods

Session 4: Scalability and Topic Models

- 4.1. Student-led discussion(s): GPs and Factorization
- 4.2. Scaling Inference
- 4.3. Topic Models

Session 5: Dealing with Unstructured Data

- 5.1. Student-led discussion(s): SHOPPER and MMMs for Choice
- 5.2. Embeddings
- 5.3. Generative Models

Session 6: Adaptive and Reinforcement Learning

- 6.1. Student-led discussion(s): Unstructured data
- 6.2. Basics of bandits, Bayesian optimization, and RL
- 6.3. Connections to dynamic choice models