MKTG 476/776, STAT 476:
Applied Probability Models in Marketing

Spring 2010

Professor Peter S. Fader
faderp@wharton.upenn.edu

Monday 3-6PM (grads)
Wednesday 3-6PM (undergrads)

Teaching assistant: Eric Schwartz
Office hours: TBA

Motivations and Objectives

Over the past five decades, statisticians have developed a number of models that have proven to be highly effective in their ability to explain and predict empirical patterns within many areas in business and the social sciences. These models use some basic building blocks from probability theory to offer behaviorally plausible perspectives on different types of timing, counting, and choice processes. Researchers in marketing have actively contributed to (and benefited from) these models for a wide variety of applications, such as new product sales forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques.

As new forms of information technology provide increasingly rich descriptions of individual-level shopping/purchasing behavior, these models offer great value to practicing managers, particularly those interested in pursuing CRM ("customer relationship management") activities. Furthermore, as more managers become comfortable with non-linear optimization techniques (using, for example, the "Solver" feature within Microsoft Excel), the specification and interpretation of these models can become a regular part of the sophisticated manager's toolkit. Taken as a whole, the methodological approaches covered in this course are well-suited to address the types of questions that are being asked with increasing frequency and interest by investors and managers of today's data-intensive businesses.

The principal objectives of this course are:

- To familiarize students with probability models and their role in marketing, information systems, and other related areas,

- To provide students with the analytical and empirical skills required to develop probability models and apply them to problems of genuine managerial interest.

Prerequisites

This course is open to students at any level (undergraduate, MBA, PhD) who have sufficient mathematical skills to handle the advanced methods that will be introduced and discussed here. It is essential that students be familiar with basic integral calculus. Furthermore, a mid-level probability/statistics course would be helpful. But aptitude to learn and fully understand this type of material is more important than mere exposure to it.

Course Organization and Materials

Every session will be lecture-based, with a strong emphasis on real-time problem solving, including mathematical derivations and numerical investigations using Microsoft Excel. Central to the development of the skills associated with probability modeling is hands-on experience. To this end, a set of homework exercises will be assigned for most sessions.

There is no formal textbook for the course (since no suitable book exists), but lecture notes covering most of the material presented in class will be made available immediately after each session. Some (but not all) of the Excel spreadsheets used in class will be made available to the students, and some journal articles will be suggested as illustrations/applications of the techniques discussed. While it is expected that students will read and review all of these materials thoroughly, there will be no pre-class readings assigned for most sessions.

Teaching Approach

The methods covered in this course will be largely unfamiliar to most students. As such, it is important to ensure that the first exposure is impactful and that there are opportunities to work with the materials multiple times and through multiple formats. To address these issues we will utilize a "heads up" learning system in the classroom. The basic elements include:

- Mandatory classroom attendance
- No laptops will be permitted in the classroom
- Presentation decks will not be provided until after class
- Each session will be recorded and made available to students (in a rich multimedia format) immediately after class

These steps are intended to help students keep their "heads up" to focus on the main points in each session. Students will be encouraged to ask questions about key conceptual issues, managerial applications, and the overall modeling philosophy; questions about more micro issues and technical details should be withheld (and addressed by reviewing the presentation decks and recordings after class).

Students are expected to create their own complete set of class notes after attending each session and working through the decks/recordings. It is fine – in fact, encouraged – for students to collaborate on this task, but it's best for each student to create his/her own notes. Any kind of "divide and conquer" approach will be counterproductive for the student.

Evaluation

Homework Exercises (20%): These exercises will be both analytical and numerical in nature. All of the numerical work can be completed using Excel.

Class Participation (20%): While there are no formal case discussions, students are expected to be actively engaged in the lectures, including periodic "cold calls" to provide solutions for problems discussed in class.

Three term projects: The first paper (15%) will be somewhat unstructured in that students will be asked to find a specific type of dataset to analyze carefully (details will be discussed in class). The second paper (20%) will be much more structured – all students will be given a common dataset to analyze. The third paper (25%) will be more open-ended with three basic options. Students can: (1) develop and apply a new probability model to a topic/dataset of their own choosing; (2) carry out an extensive simulation exercise to explore the properties of one or more models covered in class; or (3) conduct a comprehensive review of one application area of probability models in marketing. Exact requirements and possible topics will be discussed during the term.

All relevant University of Pennsylvania policies regarding academic integrity must be followed. Students may not submit work that has been prepared by (or in conjunction with) someone else, without explicit instructor permission. Any students who in any way misrepresent somebody else's work as their own will face severe disciplinary consequences.

Class Scheduling Issues

There are two sections of the course: one for graduate students on Mondays and one for undergraduates on Wednesdays. The same basic material will be covered in each one, but students are recommended to stick with their assigned section (although they can attend the other section with advance permission from the instructor). Both sessions will be recorded each week, and students are welcome (but certainly not expected) to examine both videos if they wish to gain additional perspectives on any session.

This semester's schedule is very strange: there are two Wednesday sessions before the first Monday, and there is an extra Monday session at the end of the term. Because we want to keep both sections operating in parallel, we will use the following scheduling plan:

- Week 1 (W 1/13 only): MBA's are encouraged to attend the undergrad section. It will be recorded and all students will be responsible for the material.
- Week "1A": We will hold an optional session to review the week 1 material on Monday 1/18. On Wednesday 1/20, we will hold a "bonus" session for the undergrads (grads are invited to attend as well).
- Weeks 2-13: Normal schedule
- Week 14 (M 4/26 only): A different "bonus" session for the grad students; undergrads are invited to attend as well.

Course Schedule

Week 1 (W 1/13 only): Introduction to probability models

Motivating problem: Forecasting customer retention. Comparisons to traditional regression-based models: “curve-fitting” vs. “model-building”. Careful derivation of a parametric mixture model (the beta-geometric). Coverage of maximum likelihood estimation and the Microsoft Excel Solver tool. General discussion about the philosophy and objectives of probability modeling.

Week “1A” (M 1/18, W 1/20): Bonus session: “The Paradox of Increasing Loyalty”

Monday will be an optional review of the Week 1 material; Wednesday will be a session that further applies/extends the model covered in Week 1 to demonstrate some unusual patterns about customer retention.

Week 2 (M 1/25, W 1/27): Models for count data

Introduction to the Poisson process and its extension to the negative binomial distribution. Evaluating goodness-of-fit. Alternative estimation approaches (e.g., method of moments). Dealing with problems of limited/missing data: truncated and shifted NBD models. Generalizing the model to allow for “spikes” at 0 or 1.

Week 3 (M 2/1, W 2/3): Repeated choice processes

Choice vs. counting. The binomial distribution. The beta distribution as a mixture model. Parameter estimation and inference.

Week 4 (M 2/8, W 2/10): Timing models

Motivating problem: forecasting new product adoption. Implementing and evaluating different timing models, particularly the exponential-gamma. Dealing with grouped data and right censoring. Introducing hazard functions. Derivation and discussion of other timing models (e.g., Weibull), and the linkages among them. Exploring the interplay between timing and counting processes.

Week 5 (M 2/15, W 2/17): Empirical Bayes methods

Conditional distributions and expectations. Combining population information (“priors”) with observed data for individuals. Regression-to-the-mean.

Week 6 (M 2/22, W 2/24): Customer base analysis
Project #1 (counting model) due

Combining the basic building blocks to create integrated models to estimate customer lifetime value and related concepts.

Week 7 (M 3/1, W 3/3): Customer base analysis (cont.)

More applications…

Spring Break (M 3/8, W 3/10)

Week 8 (M 3/15, W 3/17): Introducing covariates

Poisson regression and NBD regression for counting models. Beta-logistic (and alternative approaches) for choice models. Proportional hazard methods and covariate effects for timing models. General discussion about the different role of covariates from the perspective of an econometrician vis-à-vis a probability modeler. Applications.

Week 9 (M 3/22, W 3/24): Finite mixture and latent class methods

Looking at non-parametric (discrete) approaches to capturing heterogeneity. Interpreting support points versus cluster characteristics. Estimation issues. Overview of selection criteria for non-nested models.

Week 10 (M 3/29, W 3/31): Multi-item choice models

The multinomial choice process and the Dirichlet mixing distribution. Interplay between the beta and Dirichlet distributions. Discussion of Ehrenberg's "empirical laws."

Week 11 (M 4/5, W 4/7): Integrated models
Project #2 (covariate model) due

Combined models of counting, timing, and/or choice. Particular focus on the BB/NBD.

Week 12 (M 4/12, W 4/14): Nonstationary processes

Overview and comparison of techniques such as renewal processes, learning models, hidden Markov methods, and other approaches to capture dynamics over time

Week 13 (M 4/19, W 4/21): Panic time!

No scheduled class – time reserved for office hours to discuss term projects

Week 14 (M 4/26 only): Applications and wrap-up
Project #3 due

More applications/extensions/speculations, etc. Discussion of term projects.