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

Spring 2007

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

Class: Tuesday 3-6PM
Office hours: By appointment, JMHH 749

Teaching assistant: Kinshuk Jerath (kinshuk@wharton.upenn.edu)
Office hours: TBA

Motivations and Objectives

Over the past four 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 background to handle the advanced methods that will be introduced and discussed here. It is essential that students be very familiar with basic integral calculus. Furthermore, a mid-level probability/statistics course (such as STAT 430) would be very helpful. But aptitude to learn and fully understand this type of material is more important than mere exposure to it.

Course Organization and Materials

Most of the classes will be lecture-based, with a strong emphasis on real-time problem solving, including analytical exercises on the chalkboard 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 distributed on a session-to-session basis. Excel spreadsheets used in class will be made available to the students, and some journal articles will be suggested as illustrations/applications of some 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.

Evaluation

Homework Exercises (20%): These exercises will be both analytical and numerical in nature. All of the numerical work can be completed using Excel (although students are welcome to use other software packages if they wish).

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: Two of these papers (15% each) will be somewhat structured in that students will be asked to find specific types of datasets to analyze carefully (details will be discussed in class). The third paper (30%) 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.

Tentative Course Schedule

Week 1 (1/9): 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 2 (1/16): Models for count data

Introduction to the Poisson process 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 (1/23): Repeated choice processes

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

Week 4 (1/30): 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 (2/6): Empirical Bayes methods

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

Week 6 (2/13): Customer base analysis
Term paper #1 (counting model) due

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

Week 7 (2/20): Introducing covariates

Poisson regression and NBD regression for counting models. Beta-logistic (and alternative approaches) for choice models. General discussion about the different role of covariates for an econometrician vis-à-vis a probability modeler.

Tentative Course Schedule (cont.)

Week 8 (2/27): More fun with covariates

Proportional hazard methods and covariate effects for timing models. Applications. Getting ready for spring break.

Week 9 (3/13): 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 (3/20): Multi-item choice models
Term paper #2 (covariate model) due

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

Week 11 (3/27): Integrated models

Combined models of counting, timing, and/or choice. Particular focus on the BB/NBD, the Pareto/NBD, and the Polya-Aeppli models. Understanding the role of probability models in the Passover seder.

Week 12 (4/3): 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 (4/10): Applications galore

Quite a menu to choose from: Forecasting visiting and conversion behavior at Amazon.com; modeling acquisition, retention, and churn for a portfolio of Comcast services; the "Eskin" model of depth-of-repeat; and others to consider as well…

Week 14 (4/17): Applications and wrap-up
Term paper #3 due
More applications/extensions/speculations, etc. Discussion of term projects.