MKTG 396/896: Probability Models in Marketing and E-Commerce

Professor Peter S. Fader
faderp@wharton.upenn.edu
215.898.1132

University of Pennsylvania
Fall 2002
W 3-6PM

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 data sources from the Internet continue to grow, both in volume and in quality, these models are also becoming very relevant for e-commerce applications. While some pundits claim that the Internet has the potential to revolutionize certain aspects of buyer behavior, there is ample reason to believe that the same basic building blocks will continue to apply in the future. Many of 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 online businesses.

The objectives of this course are:

- to familiarize students with probability models and their role in marketing, e-commerce, 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. For undergraduates, this would mean a high comfort level with the kind of material covered in the usual Math 140-141 sequence.

Students would also benefit from (but are not required to have) recent exposure to a probability/statistics course. Aptitude to learn this type of material is more important than rote memorization of it.

MKTG 396/896:
Probability Models in Marketing and E-Commerce

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), although an annotated list of relevant reference books will be handed out in class. 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 handed out as illustrations/applications of some of the techniques discussed. While it is expected that students will read and review all of these handouts 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.

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. Any students who in any way misrepresent somebody else's work as their own will face severe disciplinary consequences.

MKTG 396/896:
Probability Models in Marketing and E-Commerce

Tentative Course Schedule

Week 1 (9/11): Introduction to counting models

Motivating problem: forecasting media exposure. Introduction to the Poisson process. Coverage of maximum likelihood estimation and the Microsoft Excel Solver tool. Evaluating goodness-of-fit. Introduction to heterogeneity. Deriving the negative binomial distribution, and extending it from the unit time interval to periods of arbitrary length.

Week 2 (9/18): Counting models II

Alternative estimation approaches (method of moments, "means & zeros"). Dealing with problems of limited/missing data: truncated and shifted NBD models. Generalizing the model to allow for "spikes" at 0 or 1. Coverage of other counting models (e.g., beta-geometric).

Week 3 (9/25): Zero-order choice processes

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

Week 4 (10/2): 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. Tradeoffs between "curve-fitting" vs. "model-building." Issues in judging model performance (e.g., fit, forecasting, interpretability of parameters).

Week 5 (10/9): Timing models II
Term paper #1 (counting model) due

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 6 (10/16): 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.

MKTG 396/896:
Probability Models in Marketing and E-Commerce

Week 7 (10/23): Empirical Bayes methods

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

Week 8 (10/30): Introducing covariates

Proportional hazard methods and covariate effects for timing models. Poisson regression and NBD regression for counting models. Beta-logistic (and alternative approaches) for choice models. Applications.

Week 9 (11/6): Introduction to MATLAB
Term paper #2 (covariate model) due

Revisiting NBD regression, and estimating the model using a popular matrix-oriented programming language. Deriving and estimating standard errors. Discussion of gradients, the Hessian matrix, and their role in parameter estimation.

Week 10 (11/13): 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 (11/20): Integrated models

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

Week 12 (11/27): Applications galore

Modeling repeat purchasing at CDNOW. Forecasting visiting and conversion behavior at Amazon.com. The "Eskin" model of depth-of-repeat. Roast turkey with all the trimmings. Pumpkin pie. Schmoozing with old friends and family.

Week 13 (12/4): Nonstationary processes
Term paper #3 due

"Nothing is stationary. Everything wiggles." – John Gierach, *Signs of Life*