

Applied Probability Models in Marketing Research: Introduction

Bruce G. S. Hardie
London Business School
bhardie@london.edu

Peter S. Fader
University of Pennsylvania
faderp@wharton.upenn.edu

12th Annual Advanced Research Techniques Forum
June 24-27, 2001

©2001 Bruce G. S. Hardie and Peter S. Fader

1

Problem 1: Predicting New Product Trial

(Modeling Timing Data)

2

Background

Ace Snackfoods, Inc. has developed a new snack product called Krunchy Bits. Before deciding whether or not to “go national” with the new product, the marketing manager for Krunchy Bits has decided to commission a year-long test market using IRI’s BehaviorScan service, with a view to getting a clearer picture of the product’s potential.

The product has now been under test for 24 weeks. On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.)

The marketing manager for Krunchy Bits would like a forecast of the product’s year-end performance in the test market. First, she wants a forecast of the percentage of households that will have made a trial purchase by week 52.

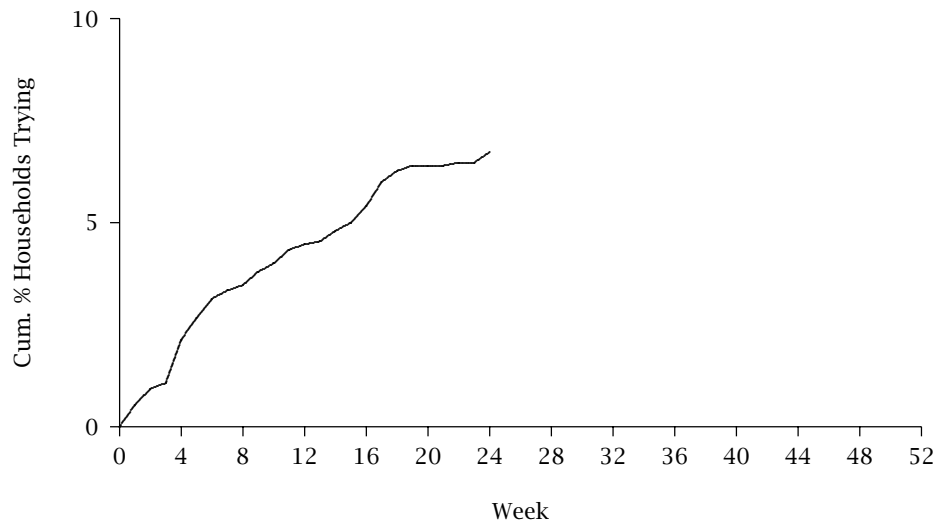
3

Krunchy Bits Cumulative Trial

Week	# Households	Week	# Households
1	8	13	68
2	14	14	72
3	16	15	75
4	32	16	81
5	40	17	90
6	47	18	94
7	50	19	96
8	52	20	96
9	57	21	96
10	60	22	97
11	65	23	97
12	67	24	101

4

Krunchy Bits Cumulative Trial



5

Approaches to Forecasting Trial

- French curve
- “Curve fitting” — specify a flexible functional form, fit it to the data, and project into the future.
- Probability model

6

Developing a Model of Trial Purchasing

- Start at the individual-level then aggregate.
 - Q:** What is the individual-level behavior of interest?
 - A:** Time (since new product launch) of trial purchase.
- We don't know exactly what is driving the behavior
⇒ treat it as a random variable.

7

The Individual-Level Model

- Let T denote the random variable of interest, and t denote a particular realization.
- Assume time-to-trial is distributed exponentially.
- The probability that an individual has tried by time t is given by:

$$F(t) = P(T \leq t) = 1 - e^{-\lambda t}$$

- λ represents the individual's trial rate.

8

The Market-Level Model

Assume two segments of consumers:

Segment	Description	Size	λ
1	ever triers	p	θ
2	never triers	$1 - p$	0

$$\begin{aligned}
 P(T \leq t) &= P(T \leq t | \text{ever trier}) \times P(\text{ever trier}) + \\
 &\quad P(T \leq t | \text{never trier}) \times P(\text{never trier}) \\
 &= pF(t | \lambda = \theta) + (1 - p)F(t | \lambda = 0) \\
 &= p(1 - e^{-\theta t})
 \end{aligned}$$

→ the “exponential w/ never triers” model

9

Estimating Model Parameters

The log-likelihood function is defined as:

$$\begin{aligned}
 LL(p, \theta | \text{data}) &= 8 \times \ln[P(0 < T \leq 1)] && + \\
 &\quad 6 \times \ln[P(1 < T \leq 2)] && + \\
 &\quad \dots && + \\
 &\quad 4 \times \ln[P(23 < T \leq 24)] && + \\
 &\quad (1499 - 101) \times \ln[P(T > 24)]
 \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -680.9$, which occurs at $\hat{p} = 0.085$ and $\hat{\theta} = 0.066$.

Forecasting Trial

- $F(t)$ represents the probability that a randomly chosen household has made a trial purchase by time t , where $t = 0$ corresponds to the launch of the new product.
- Let $T(t)$ = cumulative # households that have made a trial purchase by time t :

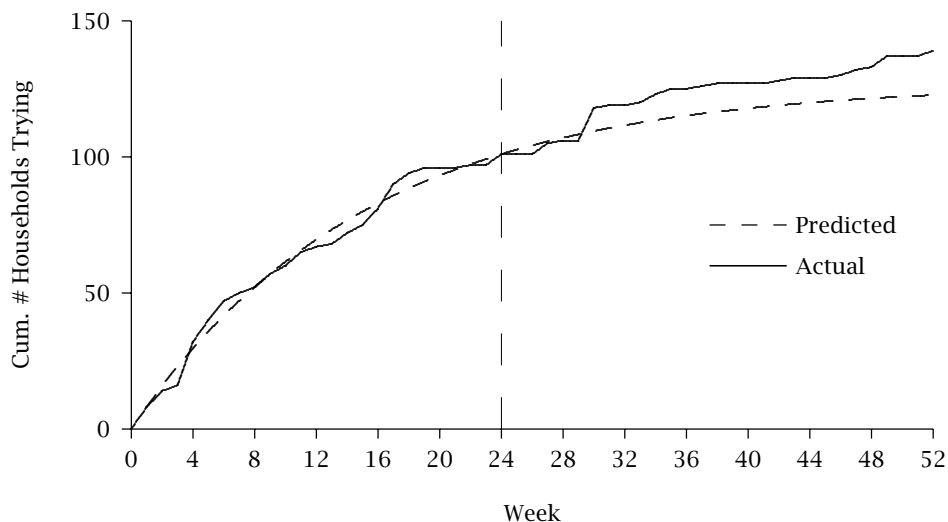
$$\begin{aligned} E[T(t)] &= N \times \hat{F}(t) \\ &= N\hat{p}(1 - e^{-\hat{\theta}t}), \quad t = 1, 2, \dots \end{aligned}$$

where N is the panel size.

- Use projection factors for market-level estimates.

11

Cumulative Trial Forecast



12

Extending the Basic Model

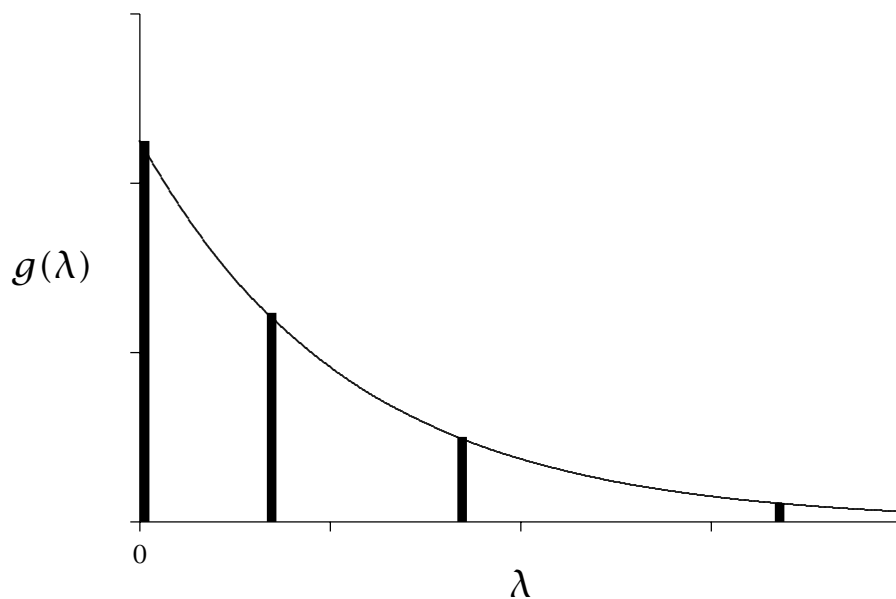
- The “exponential w/ never triers” model assumes all triers have the same underlying trial rate θ — a bit simplistic.
- Allow for multiple trier “segments” each with a different (latent) trial rate:

$$F(t) = \sum_{s=1}^S p_s F(t|\lambda_s), \quad \lambda_1 = 0, \quad \sum_{s=1}^S p_s = 1$$

- Replace the discrete distribution with a continuous distribution.

13

Distribution of Trial Rates



14

Distribution of Trial Rates

- Assume trial rates are distributed across the population according to a gamma distribution:

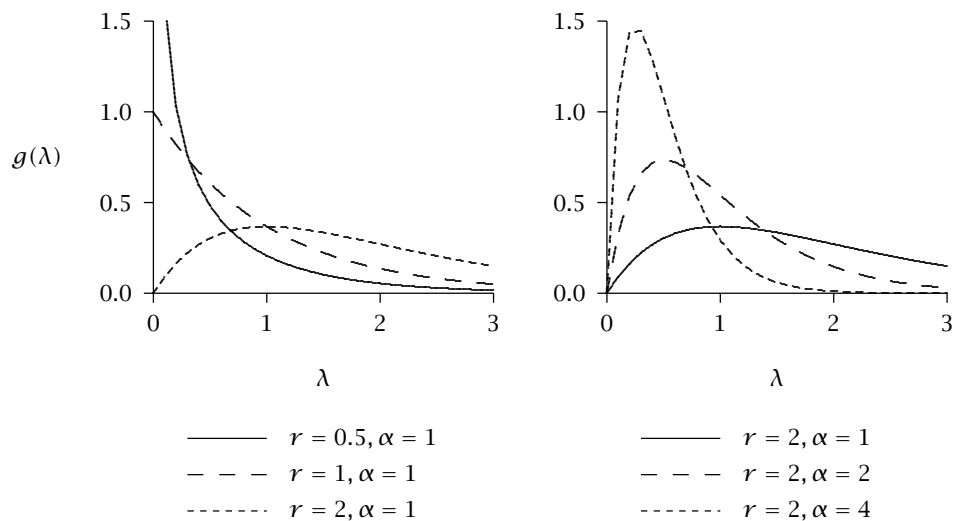
$$g(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

where r is the “shape” parameter and α is the “scale” parameter.

- The gamma distribution is a flexible (unimodal) distribution ... and is mathematically convenient.

15

Illustrative Gamma Density Functions



16

Alternative Market-Level Model

The cumulative distribution of time-to-trial at the market-level is given by:

$$\begin{aligned} P(T \leq t) &= \int_0^{\infty} P(T \leq t|\lambda) g(\lambda) d\lambda \\ &= 1 - \left(\frac{\alpha}{\alpha + t}\right)^r \end{aligned}$$

We call this the “exponential-gamma” model.

17

Estimating Model Parameters

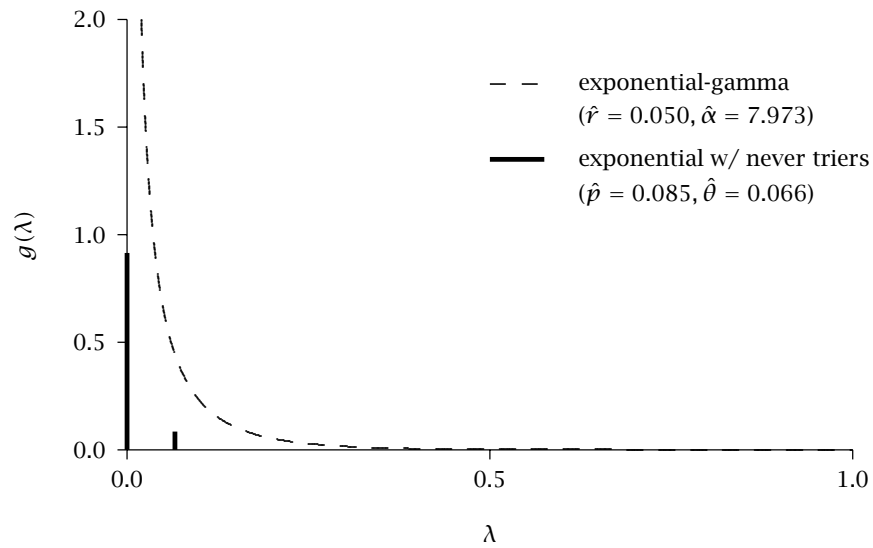
The log-likelihood function is defined as:

$$\begin{aligned} LL(r, \alpha|\text{data}) &= 8 \times \ln[P(0 < T \leq 1)] &+ \\ &6 \times \ln[P(1 < T \leq 2)] &+ \\ &\dots &+ \\ &4 \times \ln[P(23 < T \leq 24)] &+ \\ &(1499 - 101) \times \ln[P(T > 24)] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -681.4$, which occurs at $\hat{r} = 0.050$ and $\hat{\alpha} = 7.973$.

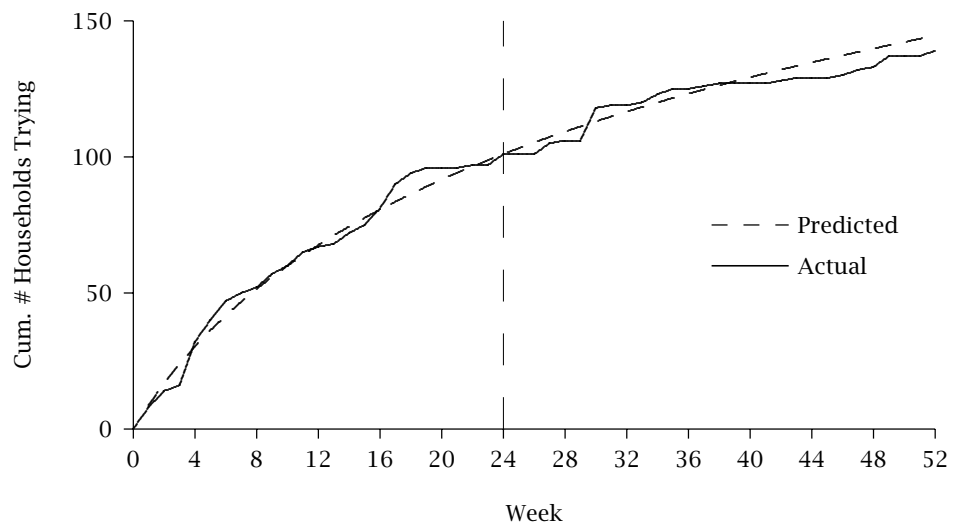
18

Estimated Distribution of λ



19

Cumulative Trial Forecast



20

Further Model Extensions

- Combine a “never triers” term with the “exponential-gamma” model.
- Incorporate the effects of marketing covariates.
- Model repeat sales using a “depth of repeat” formulation, where transitions from one repeat class to the next are modeled using an “exponential-gamma”-type model.

21

Concepts and Tools Introduced

- Probability models
- (Single-event) timing processes
- Models of new product trial/adoption

22

Further Reading

Hardie, Bruce G. S., Peter S. Fader, and Michael Wisniewski (1998), "An Empirical Comparison of New Product Trial Forecasting Models," *Journal of Forecasting*, 17 (June-July), 209-29.

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (1998), "What are the Ingredients of a 'Good' New Product Forecasting Model?" Wharton Marketing Department Working Paper #98-021.

Kalbfleisch, John D. and Ross L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, New York: Wiley.

Lawless, J.F. (1982), *Statistical Models and Methods for Lifetime Data*, New York: Wiley.

Introduction to Probability Models

The Logic of Probability Models

- Many researchers attempt to describe/predict behavior using observed variables.
- However, they still use random components in recognition that not all factors are included in the model.
- We treat behavior as if it were “random” (probabilistic, stochastic).
- We propose a model of individual-level behavior which is “summed” across individuals (taking individual differences into account) to obtain a model of aggregate behavior.

25

Uses of Probability Models

- Understanding market-level behavior patterns
- Prediction
 - To settings (e.g., time periods) beyond the observation period
 - Conditional on past behavior
- Profiling behavioral propensities of individuals
- Benchmarks/norms

26

Building a Probability Model

- (i) Determine the marketing decision problem/
information needed.
- (ii) Identify the *observable* individual-level
behavior of interest.
 - We denote this by x .
- (iii) Select a probability distribution that
characterizes this individual-level behavior.
 - This is denoted by $f(x|\theta)$.
 - We view the parameters of this distribution
as individual-level *latent traits*.

27

Building a Probability Model

- (iv) Specify a distribution to characterize the
distribution of the latent trait variable(s)
across the population.
 - We denote this by $g(\theta)$.
 - This is often called the *mixing distribution*.
- (v) Derive the corresponding *aggregate* or
observed distribution for the behavior of
interest:

$$f(x) = \int f(x|\theta)g(\theta) d\theta$$

28

Building a Probability Model

- (vi) Estimate the parameters (of the mixing distribution) by fitting the aggregate distribution to the observed data.
- (vii) Use the model to solve the marketing decision problem/provide the required information.

29

Outline

- Problem 1: Predicting New Product Trial
(Modeling Timing Data)
- Problem 2: Estimating Billboard Exposures
(Modeling Count Data)
- Problem 3: Test/Roll Decisions in Segmentation-based Direct Marketing
(Modeling “Choice” Data)
- Further applications and tools/modeling issues

30

Problem 2: Estimating Billboard Exposures

(Modeling Count Data)

31

Background

One advertising medium at the marketer's disposal is the outdoor billboard. The unit of purchase for this medium is usually a "monthly showing," which comprises a specific set of billboards carrying the advertiser's message in a given market.

The effectiveness of a monthly showing is evaluated in terms of three measures: reach, (average) frequency, and gross rating points (GRPs). These measures are determined using data collected from a sample of people in the market.

Respondents record their daily travel on maps. From each respondent's travel map, the total frequency of exposure to the showing over the survey period is counted. An "exposure" is deemed to occur each time the respondent travels by a billboard in the showing, on the street or road closest to that billboard, going towards the billboard's face.

32

Background

The standard approach to data collection requires each respondent to fill out daily travel maps for *an entire month*. The problem with this is that it is difficult and expensive to get a high proportion of respondents to do this accurately.

B&P Research is interested in developing a means by which it can generate effectiveness measures for a monthly showing from a survey in which respondents fill out travel maps for *only one week*.

Data have been collected from a sample of 250 residents who completed daily travel maps for one week. The sampling process is such that approximately one quarter of the respondents fill out travel maps during each of the four weeks in the target month.

33

Effectiveness Measures

The effectiveness of a monthly showing is evaluated in terms of three measures:

- Reach: the proportion of the population exposed to the billboard message at least once in the month.
- Average Frequency: the average number of exposures (per month) among those people reached.
- Gross Rating Points (GRPs): the mean number of exposures per 100 people.

34

Distribution of Billboard Exposures (1 week)

# Exposures	# People	# Exposures	# People
0	48	12	5
1	37	13	3
2	30	14	3
3	24	15	2
4	20	16	2
5	16	17	2
6	13	18	1
7	11	19	1
8	9	20	2
9	7	21	1
10	6	22	1
11	5	23	1

35

Modeling Objective

Develop a model that enables us to estimate a billboard showing's reach, average frequency, and GRPs for the month using the one-week data.

36

Modeling Issues

- Modeling the exposures to showing in a week.
- Estimating summary statistics of the exposure distribution for a longer period of time (i.e., one month).

37

Modeling One Week Exposures

- Let the random variable X denote the number of exposures to the showing in a week.
- At the individual-level, X is assumed to be Poisson distributed with (exposure) rate parameter λ :

$$P(X = x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$$

- Exposure rates (λ) are distributed across the population according to a gamma distribution:

$$g(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

38

Modeling One Week Exposures

- The distribution of exposures at the population-level is given by:

$$\begin{aligned} P(X = x) &= \int_0^{\infty} P(X = x|\lambda) g(\lambda) d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+1}\right)^r \left(\frac{1}{\alpha+1}\right)^x \end{aligned}$$

This is called the Negative Binomial Distribution, or NBD model.

- The mean of the NBD is given by $E(X) = r/\alpha$.

39

Computing NBD Probabilities

- Note that

$$\frac{P(X = x)}{P(X = x - 1)} = \frac{r + x - 1}{x(\alpha + 1)}$$

- We can therefore compute NBD probabilities using the following *forward recursion* formula:

$$P(X = x) = \begin{cases} \left(\frac{\alpha}{\alpha+1}\right)^r & x = 0 \\ \frac{r+x-1}{x(\alpha+1)} \times P(X = x-1) & x \geq 1 \end{cases}$$

40

Estimating Model Parameters

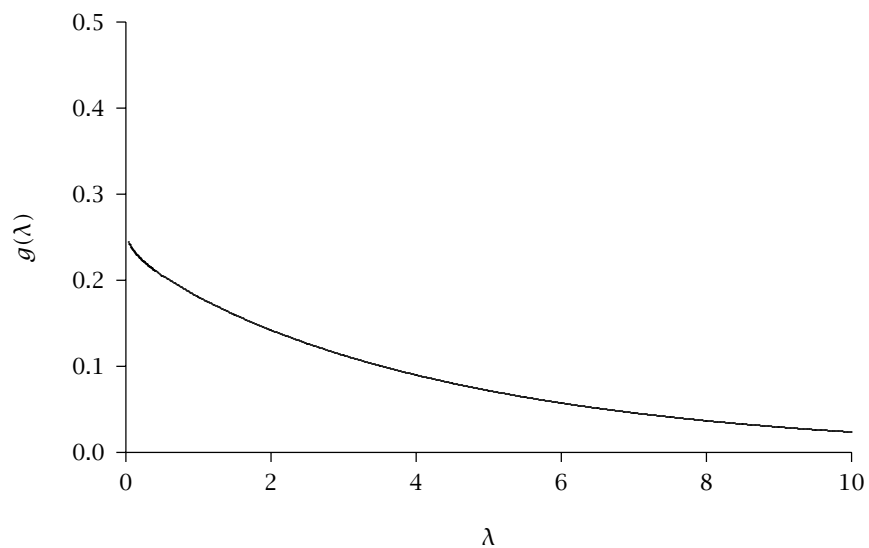
The log-likelihood function is defined as:

$$\begin{aligned} LL(r, \alpha | \text{data}) = & 48 \times \ln[P(X = 0)] + \\ & 37 \times \ln[P(X = 1)] + \\ & 30 \times \ln[P(X = 2)] + \\ & \dots + \\ & 1 \times \ln[P(X = 23)] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -649.7$, which occurs at $\hat{r} = 0.969$ and $\hat{\alpha} = 0.218$.

41

Estimated Distribution of λ



$$\hat{r} = 0.969, \hat{\alpha} = 0.218$$

42

NBD for a Non-Unit Time Period

- Let $X(t)$ be the number of exposures occurring in an observation period of length t time units.
- If, for a unit time period, the distribution of exposures *at the individual-level* is distributed Poisson with rate parameter λ , then $X(t)$ has a Poisson distribution with rate parameter λt :

$$P(X(t) = x | \lambda) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

43

NBD for a Non-Unit Time Period

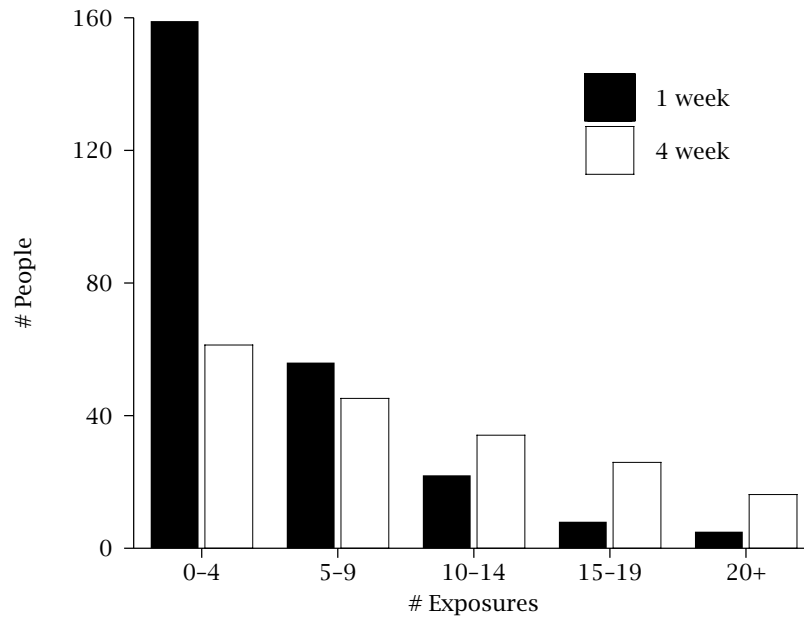
- The distribution of exposures at the population-level is given by:

$$\begin{aligned} P(X(t) = x) &= \int_0^{\infty} P(X(t) = x | \lambda) g(\lambda) d\lambda \\ &= \frac{\Gamma(r+x)}{\Gamma(r)x!} \left(\frac{\alpha}{\alpha+t}\right)^r \left(\frac{t}{\alpha+t}\right)^x \end{aligned}$$

- The mean of this distribution is given by $E[X(t)] = rt/\alpha$.

44

Exposure Distributions: 1 week vs. 4 week



45

Effectiveness of Monthly Showing

- For $t = 4$, we have:
 - $P(X(t) = 0) = 0.056$, and
 - $E[X(t)] = 17.82$
- It follows that:
 - Reach = $1 - P(X(t) = 0)$
= 94.4%
 - Frequency = $E[X(t)] / (1 - P(X(t) = 0))$
= 18.9
 - GRPs = $100 \times E[X(t)]$
= 1782

46

Concepts and Tools Introduced

- Counting processes
- The NBD model
- Extrapolating an observed histogram over time
- Using models to estimate “exposure distributions” for media vehicles

47

Further Reading

Greene, Jerome D. (1982), *Consumer Behavior Models for Non-Statisticians*, New York: Praeger.

Morrison, Donald G. and David C. Schmittlein (1988), “Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?” *Journal of Business and Economic Statistics*, **6** (April), 145-59.

Ehrenberg, A. S. C. (1988), *Repeat-Buying*, 2nd edn., London: Charles Griffin & Company, Ltd. (Available online at <http://www.empgens.com/ehrenberg.html#repeat>.)

48

Problem 3:
Test/Roll Decisions in
Segmentation-based Direct Marketing
(Modeling “Choice” Data)

49

The “Segmentation” Approach

1. Divide the customer list into a set of (homogeneous) segments.
2. Test customer response by mailing to a random sample of each segment.
3. Rollout to segments with a response rate (RR) above some cut-off point,

$$\text{e.g., } RR > \frac{\text{cost of each mailing}}{\text{unit margin}}$$

50

Ben's Knick Knacks, Inc.

- A consumer durable product (unit margin = \$161.50, mailing cost per 10,000 = \$3343)
- 126 segments formed from customer database on the basis of past purchase history information
- Test mailing to 3.24% of database

51

Ben's Knick Knacks, Inc.

Standard approach:

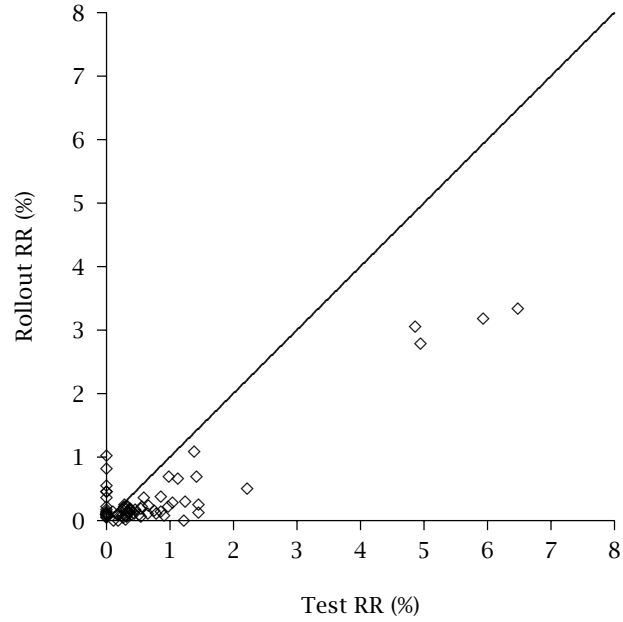
- Rollout to all segments with

$$\text{Test RR} > \frac{3343/10,000}{161.50} = 0.00207$$

- 51 segments pass this hurdle

52

Test vs. Actual Response Rate



53

Modeling Objective

Develop a model that leverages the whole data set to make better informed decisions.

Model Development

Notation:

N_s = size of segment s ($s = 1, \dots, S$)

m_s = # members of segment s tested

X_s = # responses to test in segment s

Assume: All members of segment s have the same (unknown) response probability $p_s \Rightarrow X_s$ is a binomial random variable

$$P(X_s = x_s | m_s, p_s) = \binom{m_s}{x_s} p_s^{x_s} (1 - p_s)^{m_s - x_s}$$

55

Distribution of Response Probabilities

- Heterogeneity in p_s is captured using a beta distribution:

$$g(p_s) = \frac{1}{B(\alpha, \beta)} p_s^{\alpha-1} (1 - p_s)^{\beta-1}$$

- The beta function, $B(\alpha, \beta)$, can be expressed as

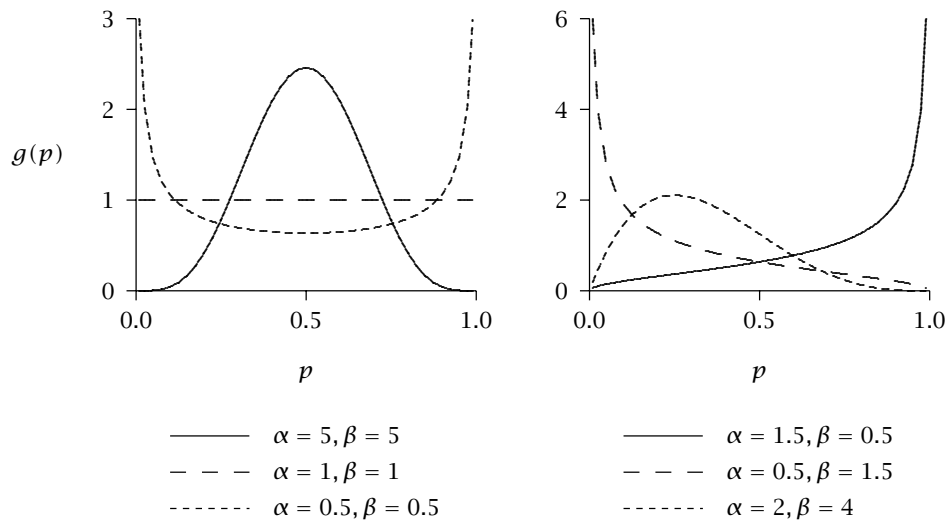
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

- The mean of the beta distribution is given by

$$E(p_s) = \frac{\alpha}{\alpha + \beta}$$

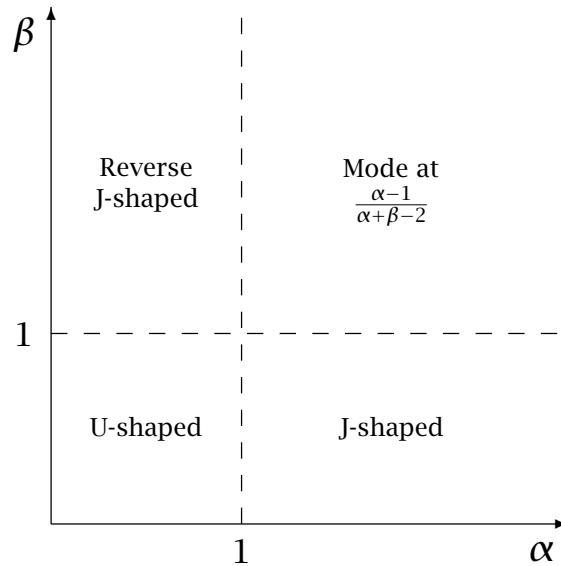
56

Illustrative Beta Density Functions



57

Shape of the Beta Density



58

The Beta Binomial Model

The aggregate distribution of responses to a mailing of size m_s is given by

$$\begin{aligned} P(X_s = x_s | m_s) &= \int_0^1 P(X_s = x_s | m_s, p_s) g(p_s) dp_s \\ &= \binom{m_s}{x_s} \frac{B(\alpha + x_s, \beta + m_s - x_s)}{B(\alpha, \beta)} \end{aligned}$$

59

Estimating Model Parameters

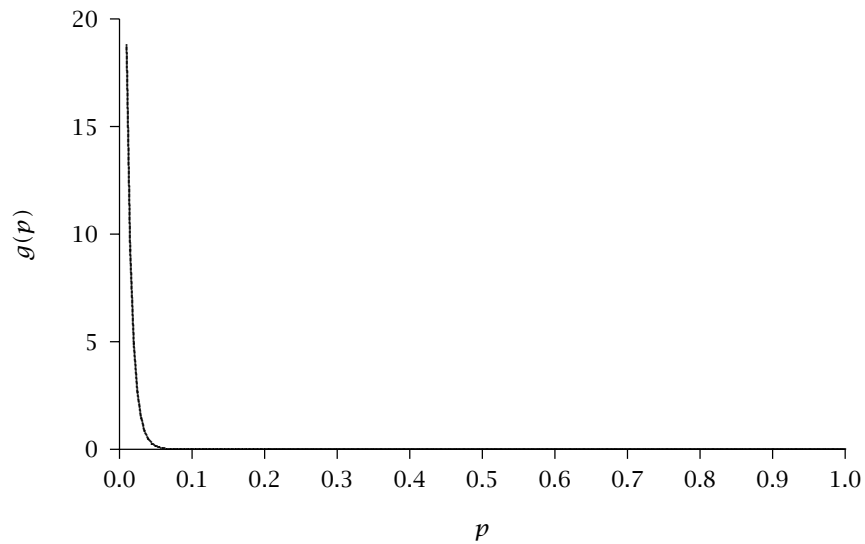
The log-likelihood function is defined as:

$$\begin{aligned} LL(\alpha, \beta | \text{data}) &= \sum_{s=1}^{126} \ln[P(X_s = x_s | m_s)] \\ &= \sum_{s=1}^{126} \ln \left[\frac{m_s!}{(m_s - x_s)! x_s!} \underbrace{\frac{\Gamma(\alpha + x_s) \Gamma(\beta + m_s - x_s)}{\Gamma(\alpha + \beta + m_s)}}_{B(\alpha + x_s, \beta + m_s - x_s)} \underbrace{\frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \Gamma(\beta)}}_{1/B(\alpha, \beta)} \right] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -200.5$, which occurs at $\hat{\alpha} = 0.439$ and $\hat{\beta} = 95.411$.

60

Estimated Distribution of p



$$\hat{\alpha} = 0.439, \hat{\beta} = 95.411, \bar{p} = 0.0046$$

61

Applying the Model

What is our best guess of p_s given a response of x_s to a test mailing of size m_s ?

Intuitively, we would expect

$$E(p_s | x_s, m_s) \approx \omega \frac{\alpha}{\alpha + \beta} + (1 - \omega) \frac{x_s}{m_s}$$

62

Bayes Theorem

- The *prior distribution* $g(p)$ captures the possible values p can take on, prior to collecting any information about the specific individual.
- The *posterior distribution* $g(p|x)$ is the conditional distribution of p , given the observed data x . It represents our updated opinion about the possible values p can take on, now that we have some information x about the specific individual.
- According to Bayes theorem:

$$g(p|x) = \frac{f(x|p)g(p)}{\int f(x|p)g(p) dp}$$

63

Bayes Theorem

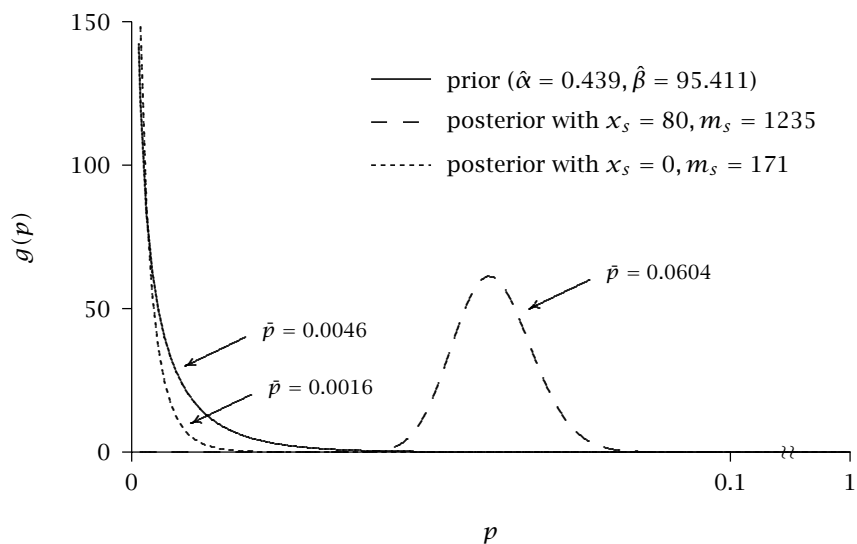
For the beta-binomial model, we have:

$$\begin{aligned}
 g(p_s|X_s = x_s, m_s) &= \frac{\overbrace{P(X_s = x_s|m_s, p_s)}^{\text{binomial}} \overbrace{g(p_s)}^{\text{beta}}}{\underbrace{\int_0^1 P(X_s = x_s|m_s, p_s) g(p_s) dp_s}_{\text{beta-binomial}}} \\
 &= \frac{1}{B(\alpha + x_s, \beta + m_s - x_s)} p_s^{\alpha+x_s-1} (1 - p_s)^{\beta+m_s-x_s-1}
 \end{aligned}$$

which is a beta distribution with parameters $\alpha + x_s$ and $\beta + m_s - x_s$.

64

Distribution of p



65

Applying the Model

Recall that the mean of the beta distribution is $\alpha/(\alpha + \beta)$. Therefore

$$E(p_s | X_s = x_s, m_s) = \frac{\alpha + x_s}{\alpha + \beta + m_s}$$

which can be written as

$$\left(\frac{\alpha + \beta}{\alpha + \beta + m_s} \right) \frac{\alpha}{\alpha + \beta} + \left(\frac{m_s}{\alpha + \beta + m_s} \right) \frac{x_s}{m_s}$$

- a weighted average of the test RR (x_s/m_s) and the population mean ($\alpha/(\alpha + \beta)$).
- “Regressing the test RR to the mean”

66

Model-Based Decision Rule

- Rollout to segments with:

$$E(p_s | X_s = x_s, m_s) > \frac{3343/10,000}{161.5} = 0.00207$$

- 66 segments pass this hurdle
- To test this model, we compare model predictions with managers' actions. (We also examine the performance of the "standard" approach.)

67

Results

	Standard	Manager	Model
# Segments (Rule)	51		66
# Segments (Act.)	46	71	53
Contacts	682,392	858,728	732,675
Responses	4,463	4,804	4,582
Profit	\$492,651	\$488,773	\$495,060

Use of model results in a profit increase of \$6287;
126,053 fewer contacts, saved for another offering.

68

Concepts and Tools Introduced

- “Choice” processes
- The Beta Binomial model
- “Regression-to-the-mean” and the use of models to capture such an effect
- Bayes theorem (and “empirical Bayes” methods)
- Using “empirical Bayes” methods in the development of targeted marketing campaigns

69

Further Reading

Colombo, Richard and Donald G. Morrison (1988), “Blacklisting Social Science Departments with Poor Ph.D. Submission Rates,” *Management Science*, **34** (June), 696–706.

Morwitz, Vicki G. and David C. Schmittlein (1998), “Testing New Direct Marketing Offerings: The Interplay of Management Judgment and Statistical Models,” *Management Science*, **44** (May), 610–28.

Sabavala, Darius J. and Donald G. Morrison (1977), “A Model of TV Show Loyalty,” *Journal of Advertising Research*, **17** (December), 35–43.

70

Further Applications and Tools/ Modeling Issues

71

Recap

- The preceding three problems introduce simple models for three behavioral processes:
 - Timing → “when”
 - Counting → “how many”
 - “Choice” → “whether/which”
- Each of these simple models has multiple applications.
- More complex behavioral phenomena can be captured by combining models from each of these processes.

72

Further Applications: Timing Models

- Repeat purchasing of new products
- Response times:
 - Coupon redemptions
 - Survey response
 - Direct mail (response, returns, repeat sales)
- Customer retention/attrition
- Other durations:
 - Salesforce job tenure
 - Length of web site browsing session

73

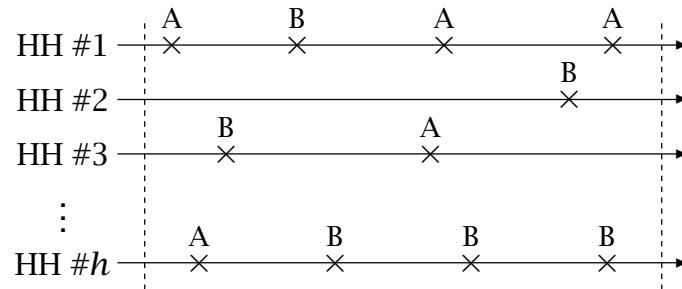
Further Applications: Count Models

- Repeat purchasing
- Customer concentration (“80/20” rules)
- Salesforce productivity/allocation
- Number of page views during a web site browsing session

74

Further Applications: “Choice” Models

- Brand choice



- Media exposure
- Multibrand choice (BB → Dirichlet Multinomial)
- Taste tests (discrimination tests)
- “Click-through” behavior

75

Integrated Models

- Counting + Timing
 - catalog purchases (purchasing | “alive” & “death” process)
 - “stickiness” (# visits & duration/visit)
- Counting + Counting
 - purchase volume (# transactions & units/transaction)
 - page views/month (# visits & pages/visit)
- Counting + Choice
 - brand purchasing (category purchasing & brand choice)
 - “conversion” behavior (# visits & buy/not-buy)

76

Further Issues

Relaxing usual assumptions:

- Non-exponential purchasing (greater regularity)
→ non-Poisson counts
- Non-gamma/beta heterogeneity (e.g., “hard core” nonbuyers, “hard core” loyals)
- Nonstationarity — latent traits vary over time

The basic models are quite robust to these departures.

77

Extensions

- Latent class/finite mixture models
- Introducing covariate effects
- Hierarchical Bayes methods

78

The Excel spreadsheets associated with this tutorial, along with electronic copies of the tutorial materials, can be found at:

<http://brucehardie.com/talks.html>