Applied Probability Models in Marketing Research: Introduction

Peter S. Fader
University of Pennsylvania
www.petefader.com

Bruce G. S. Hardie
London Business School
www.brucehardie.com

18th Annual Advanced Research Techniques Forum
June 10–13, 2007

Problem 1: Projecting Customer Retention Rates

(Modelling Discrete-Time Duration Data)

Background

One of the most important problems facing marketing managers today is the issue of *customer retention*. It is vitally important for firms to be able to anticipate the number of customers who will remain active for $1, 2, \ldots, T$ periods (e.g., years or months) after they are first acquired by the firm.

The following dataset is taken from a popular book on data mining (Berry and Linoff, *Data Mining Techniques*, Wiley 2004). It documents the "survival" pattern over a seven-year period for a sample of customer who were all "acquired" in the same period.

Customers Surviving At Least 0–7 Years

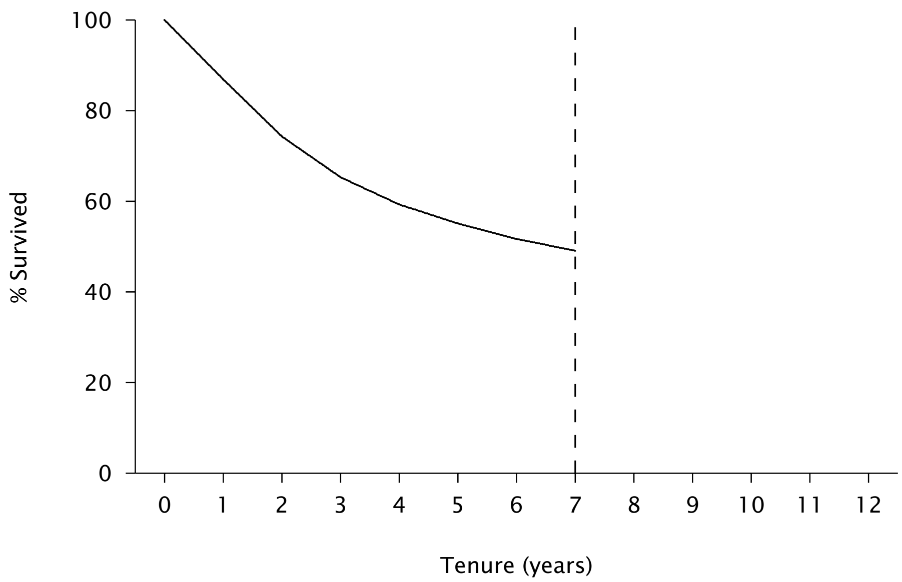
Year	# Customers	% Alive
0	1000	100.0%
1	869	86.9%
2	743	74.3%
3	653	65.3%
4	593	59.3%
5	551	55.1%
6	517	51.7%
7	491	49.1%

Of the 1000 initial customers, 869 renew their contracts at the end of the first year. At the end of the second year, 743 of these 869 customers renew their contracts.

Modelling Objective

Develop a model that enables us to project the survival curve (and therefore retention rates) over the next five years (i.e., out to $T = 12$).

Modeling Objective



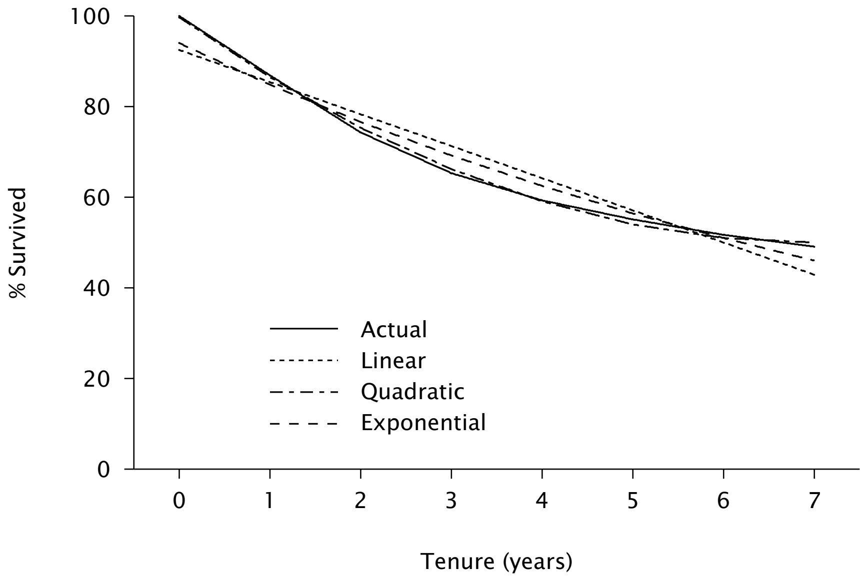
Natural Starting Point

Project survival using simple functions of time:

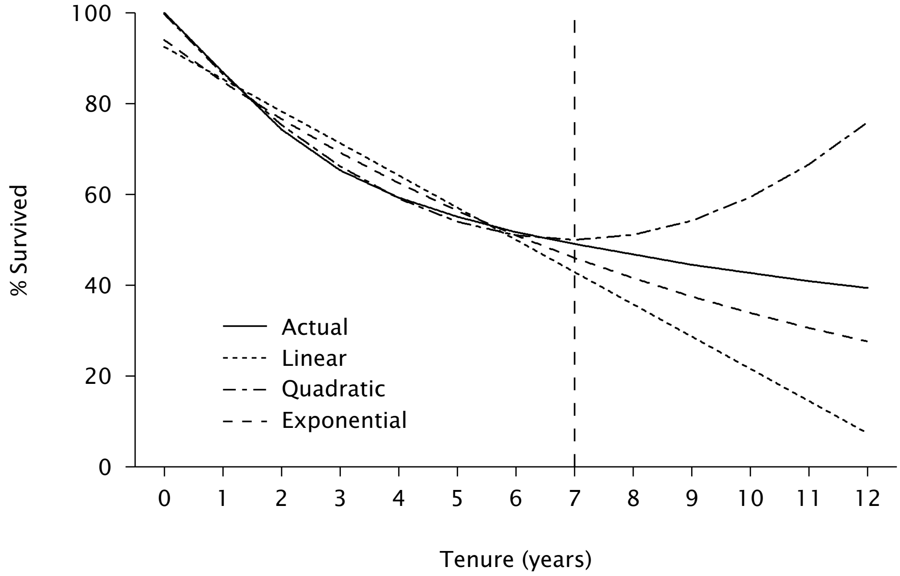
- Consider linear, quadratic, and exponential functions
- Let y = the proportion of customers surviving at least t years

$$
\begin{aligned}
y &= 0.925 - 0.071t & R^2 &= 0.922 \\
y &= 0.997 - 0.142t + 0.010t^2 & R^2 &= 0.998 \\
\ln(y) &= -0.062 - 0.102t & R^2 &= 0.964
\end{aligned}
$$

Model Fit



Survival Curve Projections



Developing a Better Model (I)

Consider the following story of customer behavior:

i. At the end of each period, an individual renews his contract with (constant and unobserved) probability $1 - \theta$.

ii. All customers have the same "churn probability" θ.

Developing a Better Model (I)

More formally:

- Let the random variable T denote the duration of the customer's relationship with the firm.
- We assume that the random variable T has a (shifted) geometric distribution with parameter θ:

$$
\begin{aligned}
P(T = t \mid \theta) &= \theta(1-\theta)^{t-1}\,, \quad t = 1, 2, 3, \ldots \\
P(T > t \mid \theta) &= (1-\theta)^{t}\,, \quad t = 1, 2, 3, \ldots
\end{aligned}
$$

Developing a Better Model (I)

The probability of the observed pattern of contract renewals is:

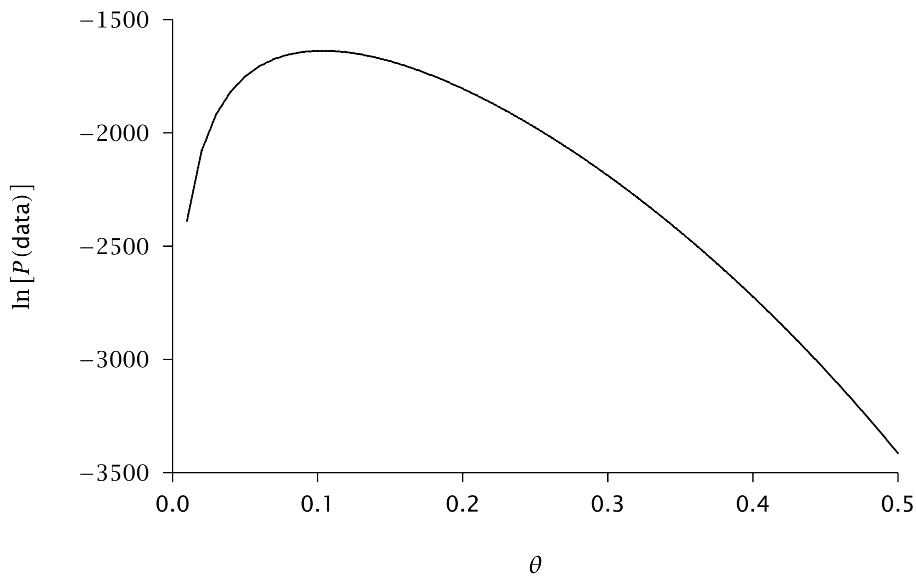
$$
\begin{aligned}
&[\theta]^{131}[\theta(1-\theta)^1]^{126}[\theta(1-\theta)^2]^{90} \\
&\quad \times [\theta(1-\theta)^3]^{60}[\theta(1-\theta)^4]^{42}[\theta(1-\theta)^5]^{34} \\
&\quad \times [\theta(1-\theta)^6]^{26}[(1-\theta)^7]^{491}
\end{aligned}
$$

Estimating Model Parameters

- Let us assume that the observed data are the outcome of a process characterized the "coin-flipping" model of contract renewal.
- Which value of θ is more likely to have "generated" the data?

θ	$P(\text{data})$	$\ln[P(\text{data})]$
0.2	1.49×10^{-784}	-1804.8
0.5	1.34×10^{-1483}	-3414.4

Estimating Model Parameters



Estimating Model Parameters

We estimate the model parameters using the method of *maximum likelihood*:

- The likelihood function is defined as the probability of observing the sample data for a given set of the (unknown) model parameters
- This probability is computed using the model and is viewed as a function of the model parameters:

$$L(\text{parameters}|\text{data}) = p(\text{data}|\text{parameters})$$

- For a given dataset, the maximum likelihood estimates of the model parameters are those values that maximize $L(\cdot)$

Estimating Model Parameters

The log-likelihood function is defined as:

$$\begin{aligned} LL(\theta|\text{data}) = {} & 131 \times \ln[P(T=1)] + \\ & 126 \times \ln[P(T=2)] + \\ & \ldots + \\ & 26 \times \ln[P(T=7)] + \\ & 491 \times \ln[P(T>7)] \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -1637.09$, which occurs at $\hat{\theta} = 0.103$.

Estimating Model Parameters

	A	B	C	D	E
1	theta	0.5000			
2	LL	-3414.44			
3					
4	Year	P(T=t)	# Cust.	# Lost	
5	0		1000		
6	1	0.5000	869	131	-90.80
7	2	0.2500	743	126	-174.67
8	3	0.1250			[illegible]7.15
9	4	0.0625	593	60	-166.36
10	5	0.0313	551	42	-145.56
11	6	0.0156	517	34	-141.40
12	7	0.0078	491	26	-126.15
13					-2382.3469
14					

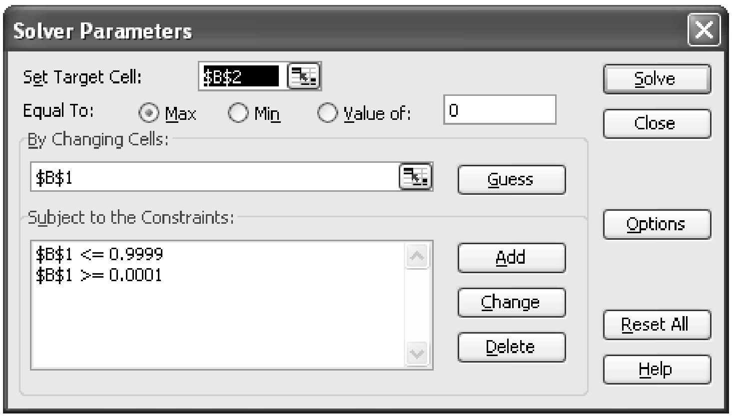
=SUM(E6:E13)

=D6*LN(B6)

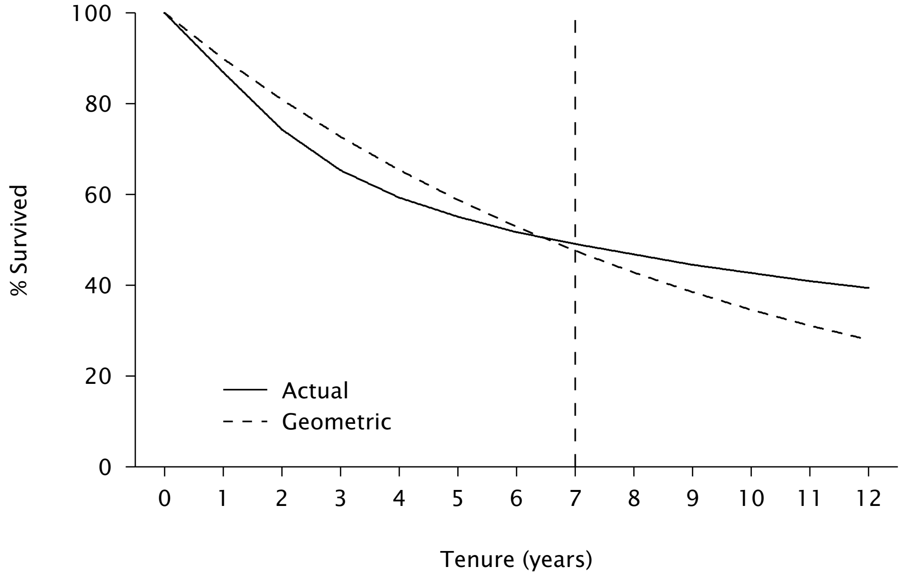
=$B$1*(1-$B$1)^(A8-1)

=C12*LN(1-SUM(B6:B12))

Estimating Model Parameters



Survival Curve Projection



What’s wrong with this story of customer contract-renewal behavior?

Developing a Better Model (II)

Consider the following story of customer behavior:

i. At the end of each period, an individual renews his contract with (constant and unobserved) probability $1 - \theta$.

ii. "Churn probabilities" vary across customers.

Developing a Better Model (II)

More formally:

- The duration of an individual customer's relationship with the firm is characterized by the (shifted) geometric distribution with parameter θ.
- Heterogeneity in θ is captured by a beta distribution with pdf

$$f(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}\,.$$

The Beta Function

- The beta function $B(a, b)$ is defined by the integral

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1}dt,\ a > 0, b > 0\,,$$

  and can be expressed in terms of gamma functions:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\,.$$

- The gamma function $\Gamma(a)$ is defined by the integral

$$\Gamma(a) = \int_0^\infty t^{a-1}e^{-t}dt,\ a > 0\,,$$

  and has the recursive property $\Gamma(a+1) = a\Gamma(a)$.

The Beta Distribution

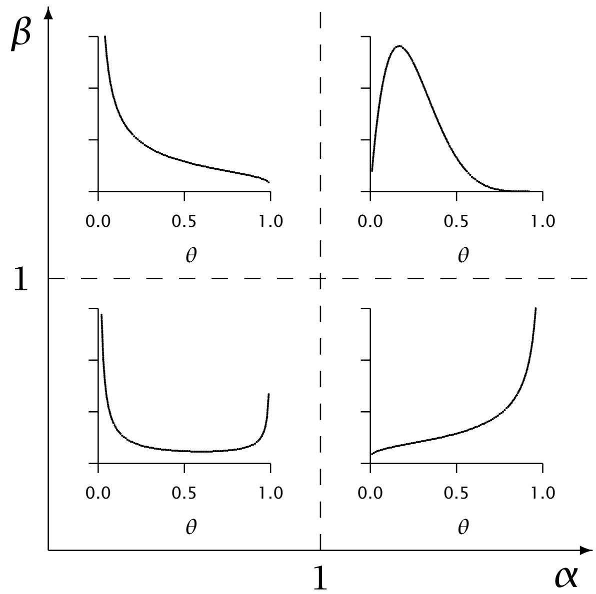
$$f(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)},\ 0 < \theta < 1\,.$$

- The mean of the beta distribution is

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

- The beta distribution is a flexible distribution … and is mathematically convenient

General Shapes of the Beta Distribution



Developing a Better Model (IIc)

For a randomly-chosen individual,

$$\begin{aligned} P(T = t \mid \alpha, \beta) &= \int_0^1 P(T = t \mid \theta) f(\theta \mid \alpha, \beta)\, d\theta \\ &= \frac{B(\alpha + 1, \beta + t - 1)}{B(\alpha, \beta)}\,. \\ P(T > t \mid \alpha, \beta) &= \int_0^1 P(T > t \mid \theta) f(\theta \mid \alpha, \beta)\, d\theta \\ &= \frac{B(\alpha, \beta + t)}{B(\alpha, \beta)}\,. \end{aligned}$$

We call this “continuous mixture” model the shifted-beta-geometric (sBG) distribution

Computing sBG Probabilities

We can compute sBG probabilities by using the following forward-recursion formula from $P(T = 1)$:

$$P(T = t) = \begin{cases} \dfrac{\alpha}{\alpha + \beta} & t = 1 \\ \\ \dfrac{\beta + t - 2}{\alpha + \beta + t - 1} P(T = t - 1) & t = 2, 3, \ldots \end{cases}$$

Estimating Model Parameters

The log-likelihood function is defined as:

$$\begin{aligned} LL(\alpha, \beta | \text{data}) = 131 \times \ln[P(T = 1)] &+ \\ 126 \times \ln[P(T = 2)] &+ \\ \ldots \qquad\qquad &+ \\ 26 \times \ln[P(T = 7)] &+ \\ 491 \times \ln[P(T > 7)] & \end{aligned}$$

The maximum value of the log-likelihood function is $LL = -1611.16$, which occurs at $\hat{\alpha} = 0.668$ and $\hat{\beta} = 3.806$.

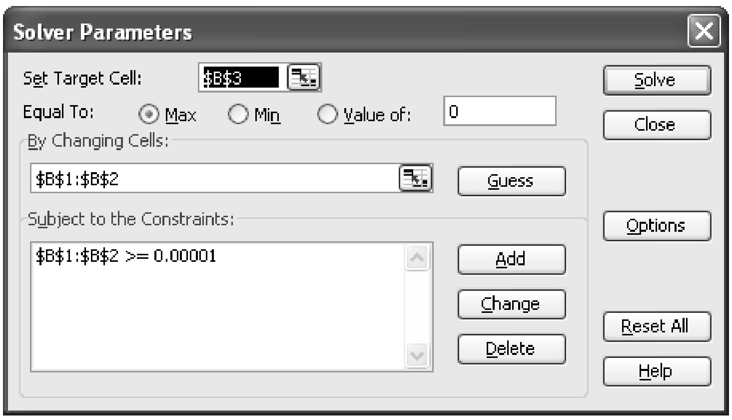
Estimating Model Parameters

	A	B	C	D	E
1	alpha	1.000			
2	beta	1.000			
3	LL	-2115.55			
4					
5	Year	P(T=t)	# Cust.	# Lost	
6	0		1000		
7	1	0.5000	[illegible]	31	-90.8023
8	2	0.1667	743	126	-225.7617
9	[illegible]	[illegible]	[illegible]	90	-223.6416
10	[illegible]	[illegible]	[illegible]	60	-179.7439
11	5	0.0333	551	42	-142.8503
12	6	0.0238	517	34	-127.0808
13	7	0.0179	491	26	-104.6591
14					-1021.0058

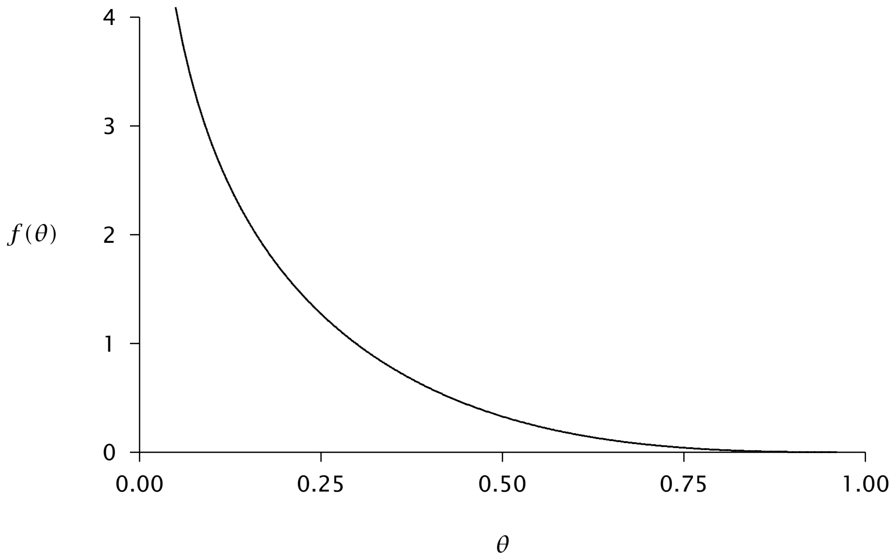
=B1/(B1+B2)

=B7*($B$2+A8-2)/($B$1+$B$2+A8-1)

Estimating Model Parameters

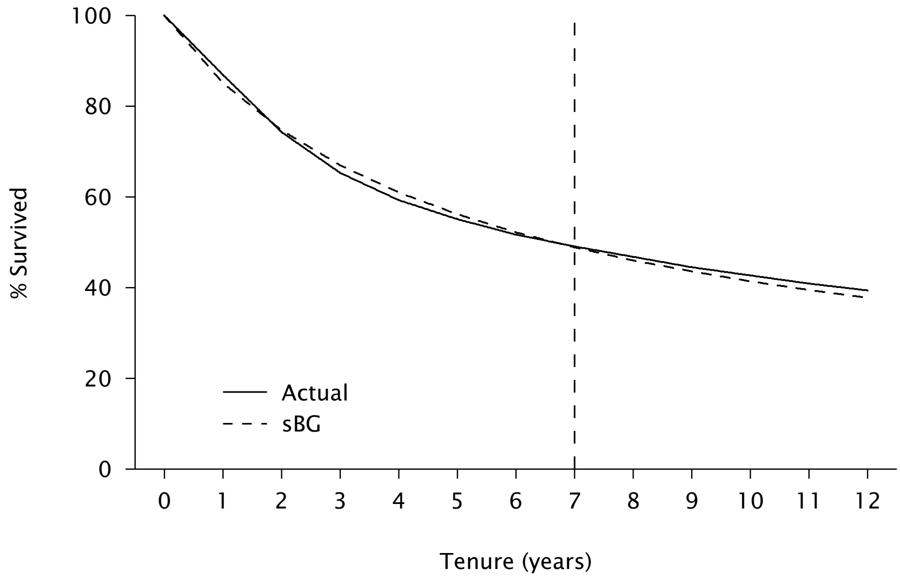


Estimated Distribution of Churn Probabilities



$$\hat{\alpha} = 0.668, \hat{\beta} = 3.806, \widehat{E(\theta)} = 0.149$$

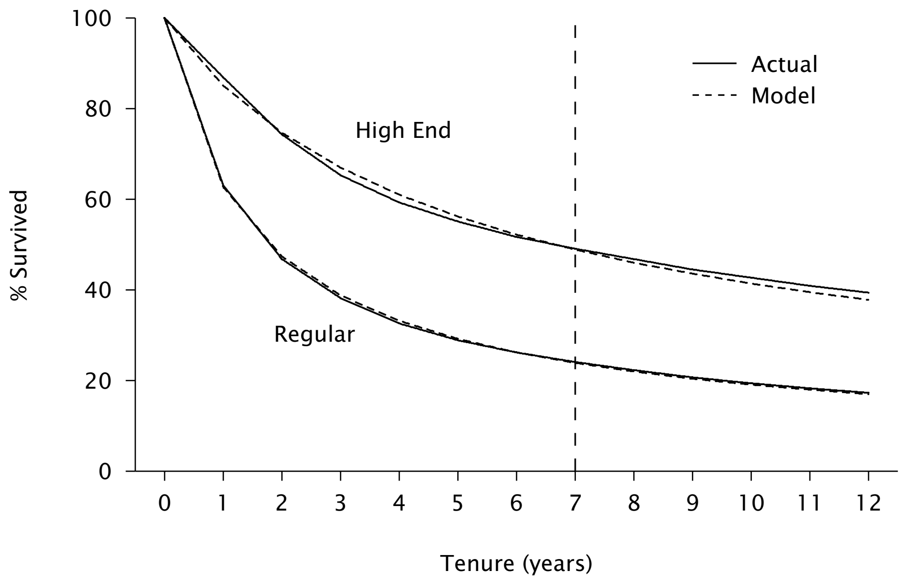
Survival Curve Projection



A Further Test of the sBG Model

- The dataset we have been analyzing is for a “high end” segment of customers.
- We also have a dataset for a “regular” customer segment.
- Fitting the sBG model to the data on contract renewals for this segment yields $\hat{\alpha} = 0.704$ and $\hat{\beta} = 1.182$ ($\Longrightarrow \widehat{E(\theta)} = 0.373$).

Survival Curve Projections



Implied Retention Rates

- The retention rate for period t (r_t) is defined as the proportion of customers who had renewed their contract at the end of period $t-1$ who then renew their contract at the end of period t.
- For any model of contract duration with survivor function $P(T>t)$,

$$r_t = \frac{P(T>t)}{P(T>t-1)}$$

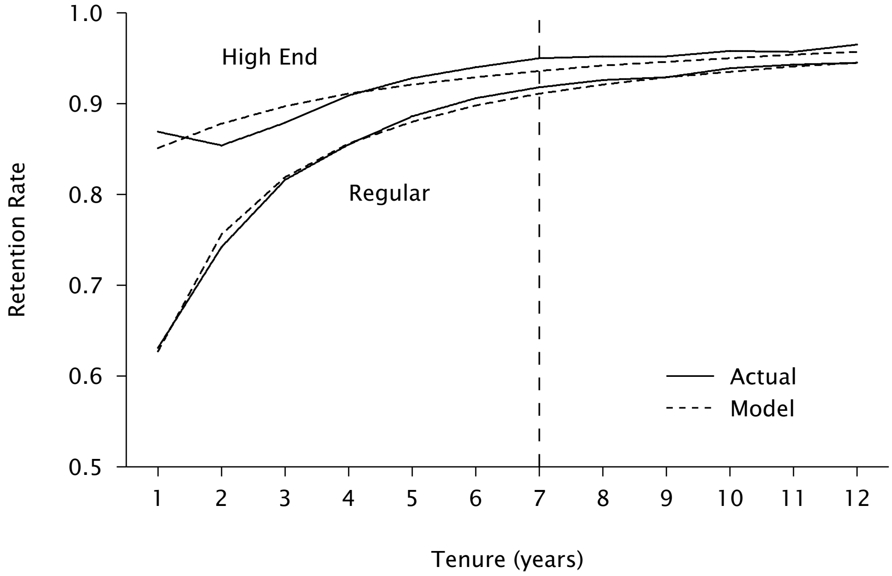
Implied Retention Rates

- For the sBG model,

$$r_t = \frac{\beta+t-1}{\alpha+\beta+t-1}$$

- An increasing function of time, even though the individual-level retention probability is constant.
- A sorting effect in a heterogeneous population.

Projecting Retention Rates



Concepts and Tools Introduced

- Probability models
- Maximum-likelihood estimation of model parameters
- Modelling discrete-time (single-event) duration data
- Models of contract renewal behavior

Further Reading

Buchanan, Bruce and Donald G. Morrison (1988), "A Stochastic Model of List Falloff with Implications for Repeat Mailings," *Journal of Direct Marketing*, **2** (Summer), 7–15.

Fader, Peter S. and Bruce G. S. Hardie (2007), "How to Project Customer Retention," *Journal of Interactive Marketing*, **21** (Winter), 76–90.

Fader, Peter S. and Bruce G. S. Hardie (2007), "How <u>Not</u> to Project Customer Retention."
<`http://brucehardie.com/notes/016/`>

Weinberg, Clarice Ring and Beth C. Gladen (1986), "The Beta-Geometric Distribution Applied to Comparative Fecundability Studies," *Biometrics*, **42** (September), 547–560.

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.

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

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 characteristics*.

Building a Probability Model

(iv) Specify a distribution to characterize the distribution of the latent characteristic 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:

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

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.

Outline

- Problem 1: Projecting Customer Retention Rates
  (Modelling Discrete-Time Duration Data)
- Problem 2: Predicting New Product Trial
  (Modelling Continuous-Time Duration Data)
- Problem 3: Estimating Billboard Exposures
  (Modelling Count Data)
- Problem 4: Test/Roll Decisions in Segmentation-based Direct Marketing
  (Modelling "Choice" Data)

Problem 2:
Predicting New Product Trial

(Modelling Continuous-Time Duration Data)

Background

Ace Snackfoods, Inc. has developed a new shelf-stable juice product called Kiwi Bubbles. Before deciding whether or not to "go national" with the new product, the marketing manager for Kiwi Bubbles 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.

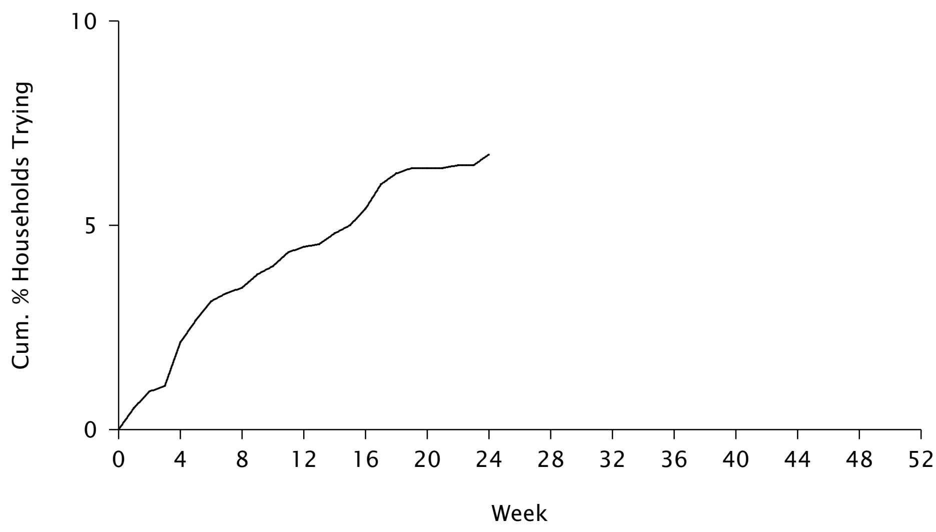
The product has now been under test for 24 weeks. On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.)

The marketing manager for Kiwi Bubbles 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.

Kiwi Bubbles 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

Kiwi Bubbles Cumulative Trial



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 $\Rightarrow$ treat it as a random variable.

The Individual-Level Model

- Let T denote the random variable of interest, and t denote a particular realization.
- Assume time-to-trial is characterized by the exponential distribution with parameter λ (which represents an individual's trial rate).
- The probability that an individual has tried by time t is given by:

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

Distribution of Trial Rates

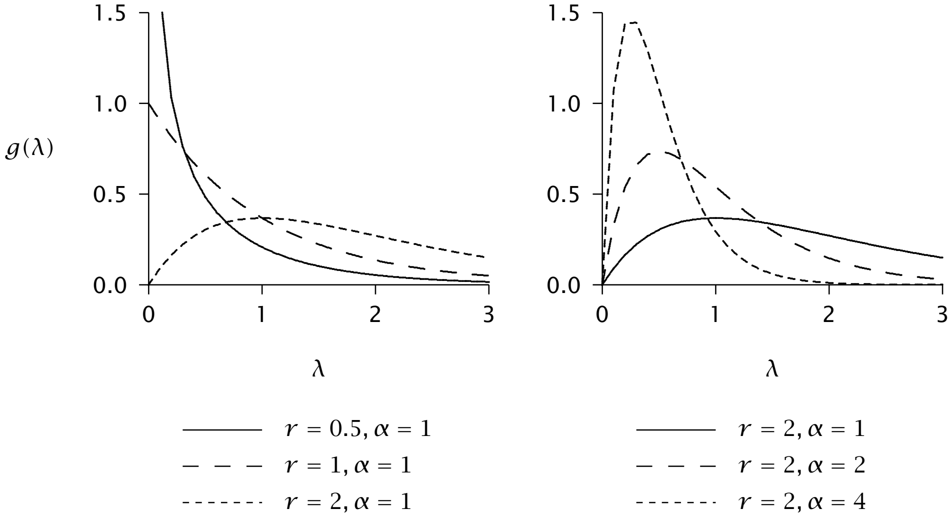
- Assume trial rates are distributed across the population according to a gamma distribution:

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

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

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

Illustrative Gamma Density Functions



Market-Level Model

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

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

We call this the “exponential-gamma” model.

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

Estimating Model Parameters

	A	B	C	D	E	F
1	Product:	Kiwi Bubbles			r	1.000
2	Panelists:	1499			alpha	1.000
3					LL	-4909.5
4		Cum_Trl				
5	Week	# HHs	Incr_Trl	P(T <= t)	P(try week t)	
6			8	0.50000	0.50000	-5.545
7	2	14	6	0.66667	0.16667	-10.751
8	3	16	2	0.75000	0.08333	-4.970
9	4	32	16	0.80000	0.05000	-47.932
10	5	40	8	0.83333		-27.210
11	6	47	7	0.85714	0.02381	-26.164
12	7	50	3	0.87500	0.01786	-12.076
13	8	52	2	0.88889	0.01389	-8.553
14	9	57	5	0.90000	0.01111	-22.499
15	10	60	3	0.90909	0.00909	-14.101
29	24	101	4	0.96000	0.00167	-25.588
30						-4499.988

=SUM(F6:F30)

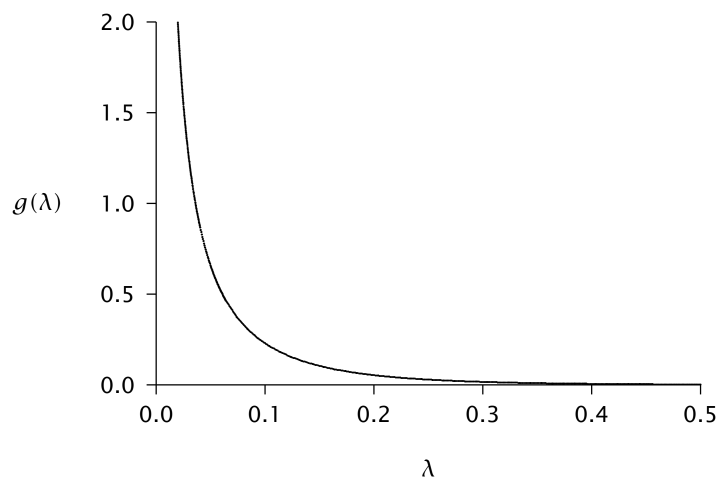
=1-(F$2/(F$2+A6))^F$1

=D7-D6

=C8*LN(E8)

=(B2-B29)*LN(1-D29)

Estimated Distribution of λ



$\hat{r} = 0.050, \hat{\alpha} = 7.973$

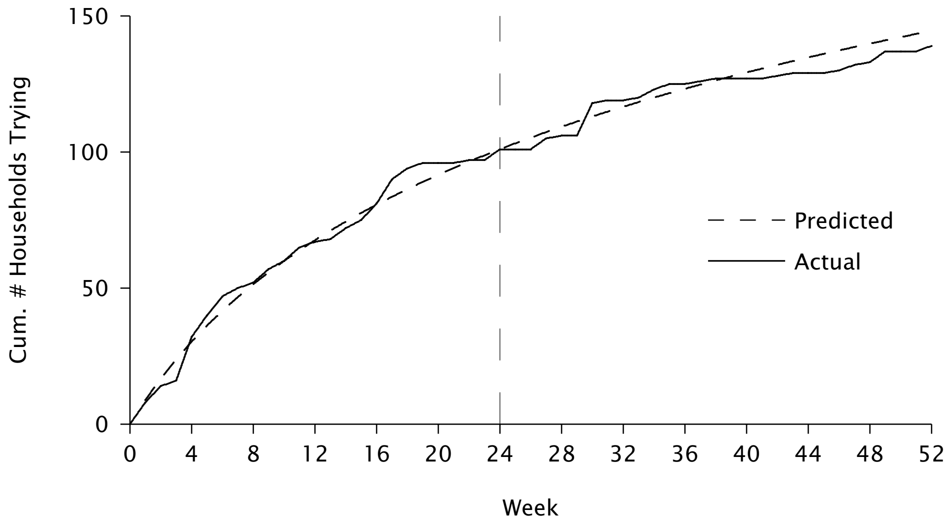
Forecasting Trial

- $F(t)$ 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:

$$\begin{aligned} E[T(t)] &= N \times \hat{F}(t) \\ &= N \left\{ 1 - \left(\frac{\hat{\alpha}}{\hat{\alpha} + t} \right)^{\hat{r}} \right\} . \end{aligned}$$

  where N is the panel size.
- Use projection factors for market-level estimates.

Cumulative Trial Forecast



Further Model Extensions

- Add a “never triers” parameter.
- 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.

Concepts and Tools Introduced

- Modelling continuous-time (single-event) duration data
- Models of new product trial

Further Reading

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (2003), "Forecasting New Product Trial in a Controlled Test Market Environment," *Journal of Forecasting*, **22** (August), 391–410.

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), 209–229.

Kalbfleisch, John D. and Ross L. Prentice (2002), *The Statistical Analysis of Failure Time Data*, 2nd edn., New York: Wiley.

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

Problem 3: Estimating Billboard Exposures

(Modelling Count Data)

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.

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.

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.

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

Average # Exposures = 4.456

Modelling 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.

Modelling Issues

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

Model Development

- 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:

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

Model Development

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

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

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

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

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:

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

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

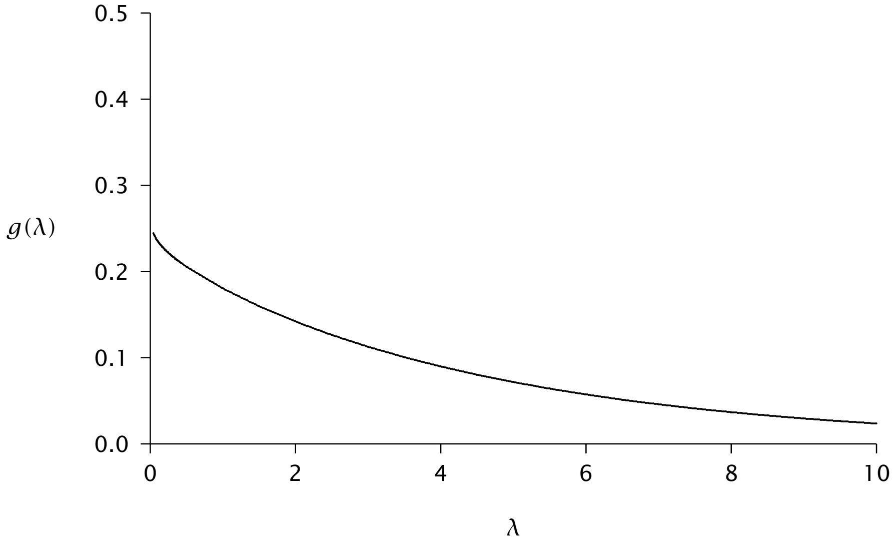
Estimating Model Parameters

	A	B	C	D
1	r	1.000		
2	alpha	1.000		
3	LL	-945.5		
4				
5	x	f_x	P(X=x)	
6	0	48	0.50000	-33.27
7	1	37	0.25000	-51.29
8	2	30	0.12500	-62.38
9	[illegible]	[illegible]	[illegible]	-66.54
10	4	20	0.03125	-69.31
11	5	16	0.01563	-66.54
12	6	13	0.00781	-63.08
13	7	11	0.00391	-61.00
14	8	9	0.00195	-56.14
15	9	7	0.00098	-48.52
28	22	1	0.00000	-15.94
29	23	1	0.00000	-16.64

=(B2/(B2+1))^B1

=C6*($B$1+A7-1)/(A7*($B$2+1))

Estimated Distribution of λ



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

NBD for a Non-Unit Time Period

- Let $X(t)$ be the number of exposures occuring 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:

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

NBD for a Non-Unit Time Period

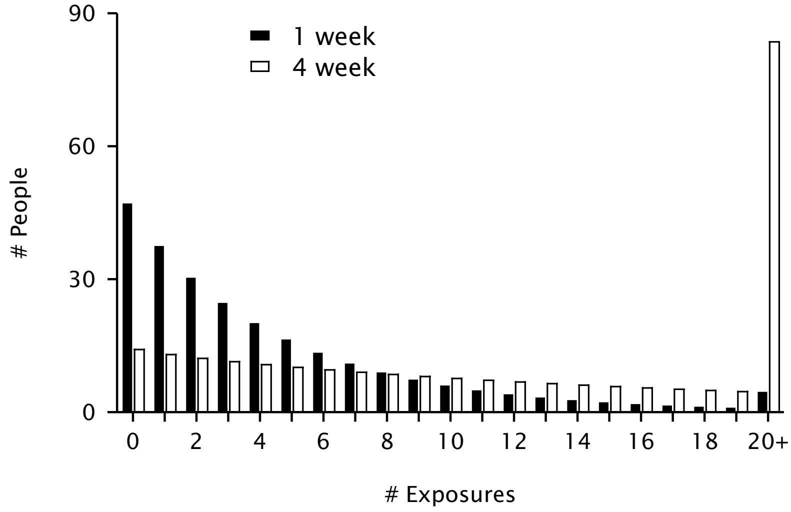
- The distribution of exposures at the population-level is given by:

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

- The mean of this distribution is given by

$$E[X(t)] = \frac{rt}{\alpha}$$

Exposure Distributions: 1 week vs. 4 week



Effectiveness of Monthly Showing

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

Concepts and Tools Introduced

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

Further Reading

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

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), 145–159.

Problem 4:
Test/Roll Decisions in Segmentation-based Direct Marketing

(Modelling "Choice" Data)

The "Segmentation" Approach

i. Divide the customer list into a set of (homogeneous) segments.

ii. Test customer response by mailing to a random sample of each segment.

iii. Rollout to segments with a response rate (RR) above some cut-off point,

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

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

Ben’s Knick Knacks, Inc.

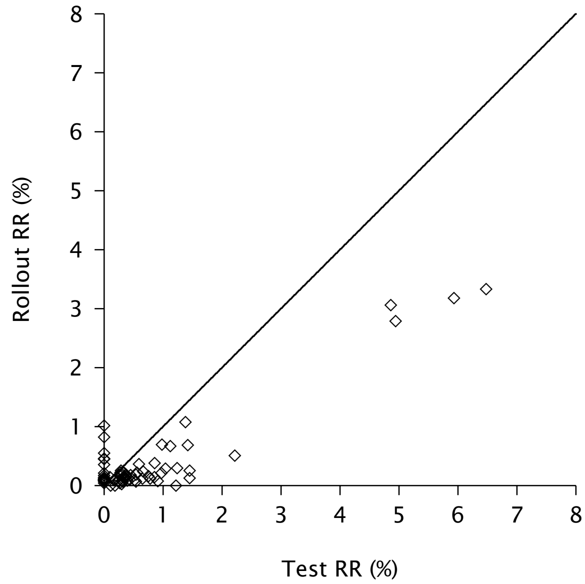
Standard approach:

- Rollout to all segments with

$$\text{Test RR} > \frac{3{,}343/10{,}000}{161.50} = 0.00207$$

- 51 segments pass this hurdle

Test vs. Actual Response Rate



Modelling Objective

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

Model Development

i. Assuming all members of segment s have the same (unknown) response probability p_s, X_s has a binomial distribution:

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

with $E(X_s|m_s, p_s) = m_s p_s$.

ii. Heterogeneity in p_s is captured using a beta distribution:

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

The Beta Binomial Model

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

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

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

Estimating Model Parameters

	A	B	C	D	E
1	alpha	1.000		B(alpha,beta)	1.000
2	beta	1.000			
3	LL	-718.9			
4					
5	Segment	m_s	x_s	P(X=x|m)	
6	1	34	0	0.02857	-3.555
7	2	102			
8	3	53			
9	4	145			
10					-7.135
11					-4.977
12					-7.120
13					
14	9	1083	24		
130	125	383	0	0.00260	-5.951
131	126	404	0	0.00247	-6.004

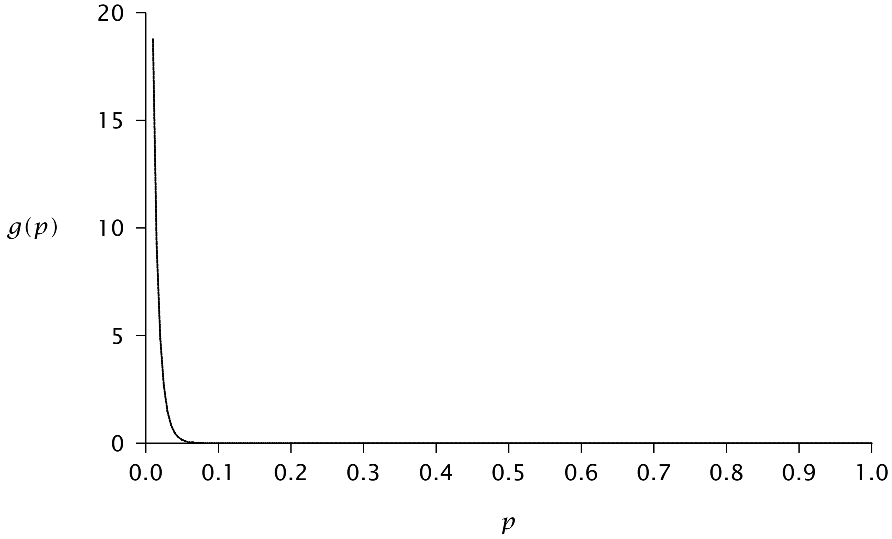
=SUM(E6:E131)

=EXP(GAMMALN(B1)+GAMMALN(B2)-GAMMALN(B1+B2))

=COMBIN(B6,C6)*EXP(GAMMALN(B$1+C6)+GAMMALN(B$2+B6-C6)-GAMMALN(B$1+B$2+B6))/E$1

=LN(D11)

Estimated Distribution of p



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

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

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

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:

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

Bayes' Theorem

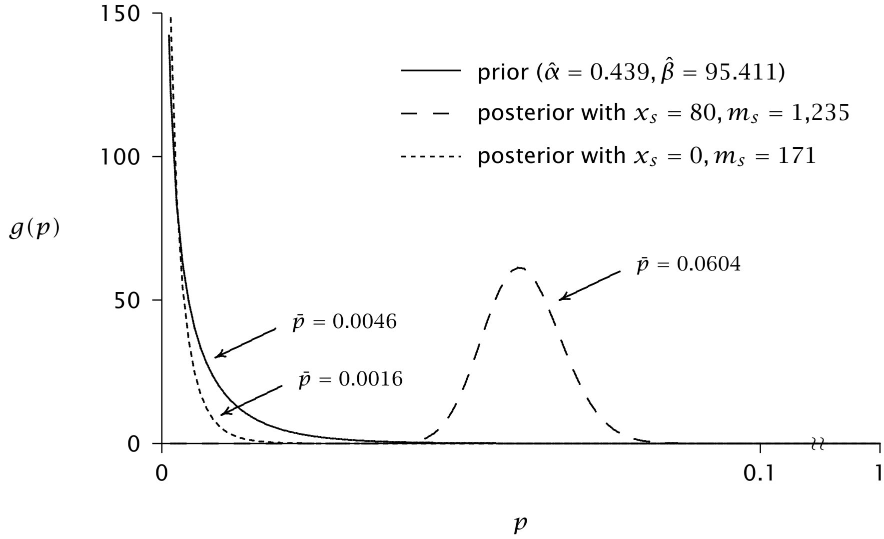
For the beta-binomial model, we have:

$$g(p_s|X_s = x_s, m_s) = \frac{\overbrace{P(X_s = x_s|m_s, p_s)}^{\text{binomial}}\,\overbrace{g(p_s)}^{\text{beta}}}{\underbrace{\int_0^1 P(X_s = x_s|m_s, p_s)\,g(p_s)\,dp_s}_{\text{beta-binomial}}}$$

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

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

Distribution of p



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"

Model-Based Decision Rule

- Rollout to segments with:

$$E(p_s|X_s = x_s, m_s) > \frac{3{,}343/10{,}000}{161.5} = 0.00207$$

- 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.)

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 $6,287; 126,053 fewer contacts, saved for another offering.

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

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.

Morrison, Donald G. and Manohar U. Kalwani (1993), "The Best NFL Field Goal Kickers: Are They Lucky or Good?" *Chance*, **6** (August), 30–37.

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), 610–628.

Discussion

Recap

- The preceding four problems introduce simple models for three behavioral processes:
  - Timing $\longrightarrow$ "when"
  - Counting $\longrightarrow$ "how many"
  - "Choice" $\longrightarrow$ "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.

Further Applications: Timing Models

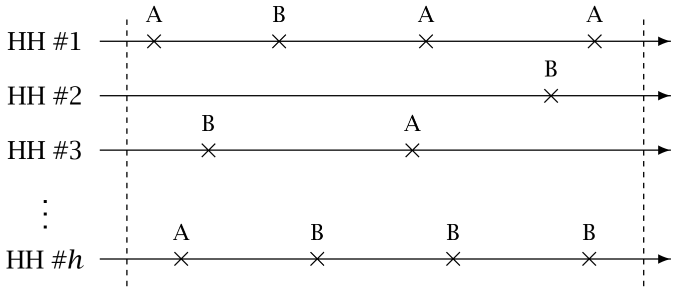
- Repeat purchasing of new products
- Response times:
  - Coupon redemptions
  - Survey response
  - Direct mail (response, returns, repeat sales)
- Other durations:
  - Salesforce job tenure
  - Length of web site browsing session

Further Applications: Count Models

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

Further Applications: "Choice" Models

- Brand choice



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

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)

A Template for Integrated Models

			Stage 2	
		Counting	Timing	Choice
	Counting			
Stage 1	Timing			
	Choice			

Further Issues

Relaxing usual assumptions:

- Non-exponential purchasing (greater regularity) $\longrightarrow$ 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.

Extensions

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

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

`http://brucehardie.com/talks.html`

An annotated list of key books for those interested in applied probability modelling can be found at:

`http://brucehardie.com/notes/001/`