An Introduction to Applied Probability Models

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

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

Workshop on Customer-Base Analysis
Johann Wolfgang Goethe-Universität, Frankfurt
March 8–9, 2006

Problem 1: Projecting Customer Retention Rates

(Modeling 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

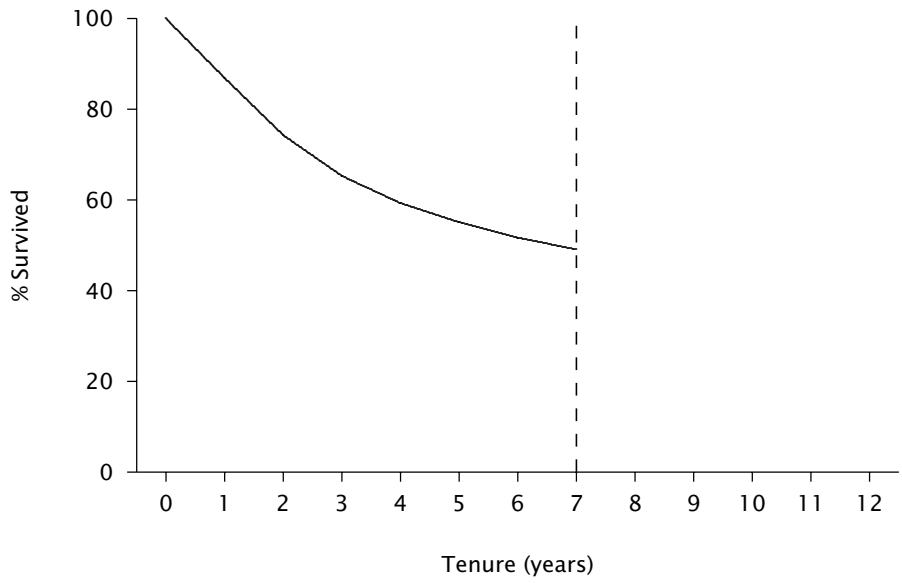
Year	# Customers	% Alive
0	1000	100.0%
1	869	86.9%
2	743	74.3%
3	653	65.3%
4	593	59.3%
5	551	55.1%
6	517	51.7%
7	491	49.1%

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.

Modeling Objective

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

Modeling Objective



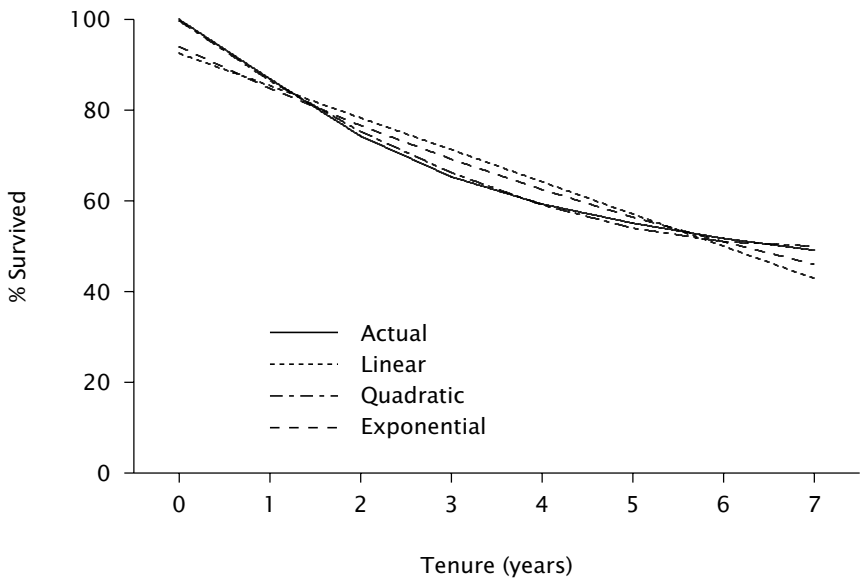
Natural Starting Point

Project survival using simple functions of time:

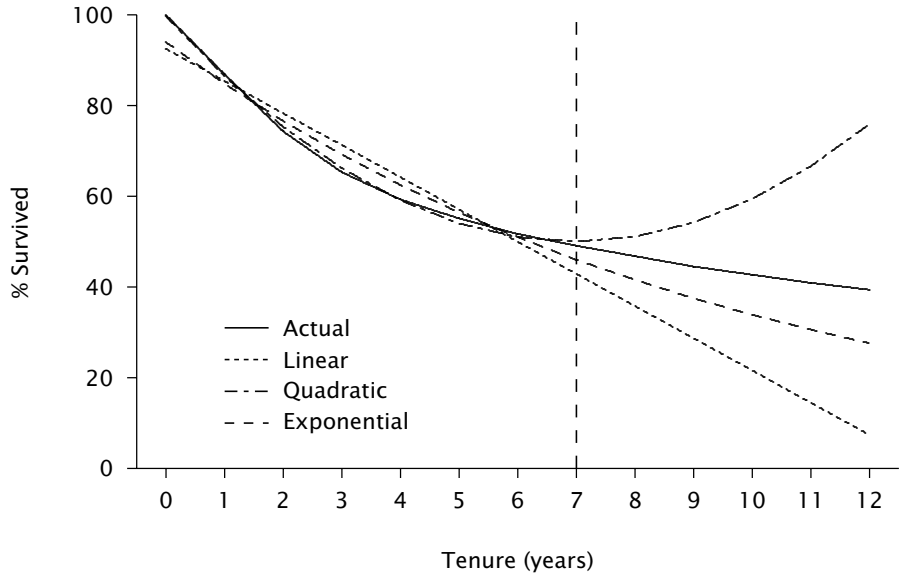
- Consider linear, quadratic, and exponential functions
- Let y = the proportion of customers surviving at least t years

$$y = 0.925 - 0.071t \qquad R^2 = 0.922$$
$$y = 0.997 - 0.142t + 0.010t^2 \qquad R^2 = 0.998$$
$$\ln(y) = -0.062 - 0.102t \qquad R^2 = 0.964$$

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:

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

Developing a Better Model (I)

The probability of the observed pattern of contract renewals is:

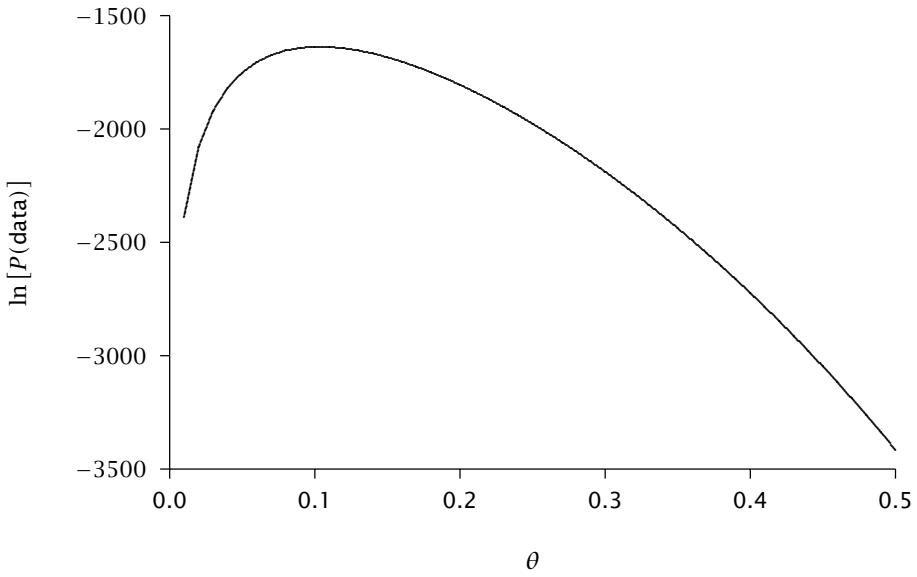
$$\begin{aligned}&[\theta]^{131}[\theta(1-\theta)^{1}]^{126}[\theta(1-\theta)^{2}]^{90}\\ &\quad\times[\theta(1-\theta)^{3}]^{60}[\theta(1-\theta)^{4}]^{42}[\theta(1-\theta)^{5}]^{34}\\ &\quad\times[\theta(1-\theta)^{6}]^{26}[(1-\theta)^{7}]^{491}\end{aligned}$$

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	9.2×10^{-797}	-1832.9
0.5	1.6×10^{-1221}	-3501.8

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 all of the data points
- This probability is computed using the model and is viewed as a function of the model parameters:

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

- For any given set of parameters, $L(\cdot)$ tells us the probability of obtaining the actual data
- 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	647	96	-137.15
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					

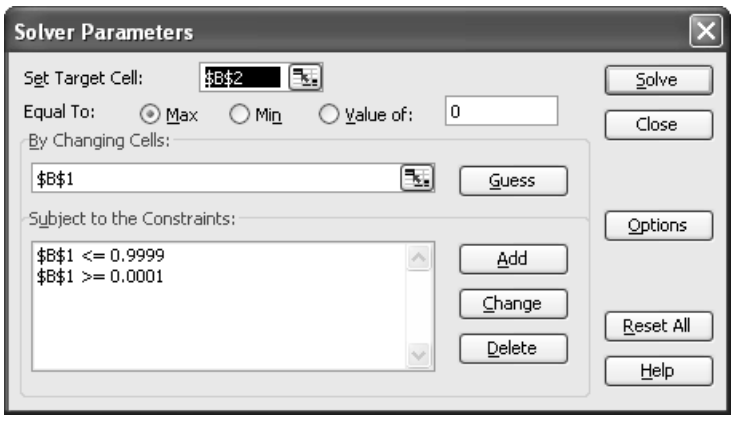
=SUM(E6:E13)

=D6*LN(B6)

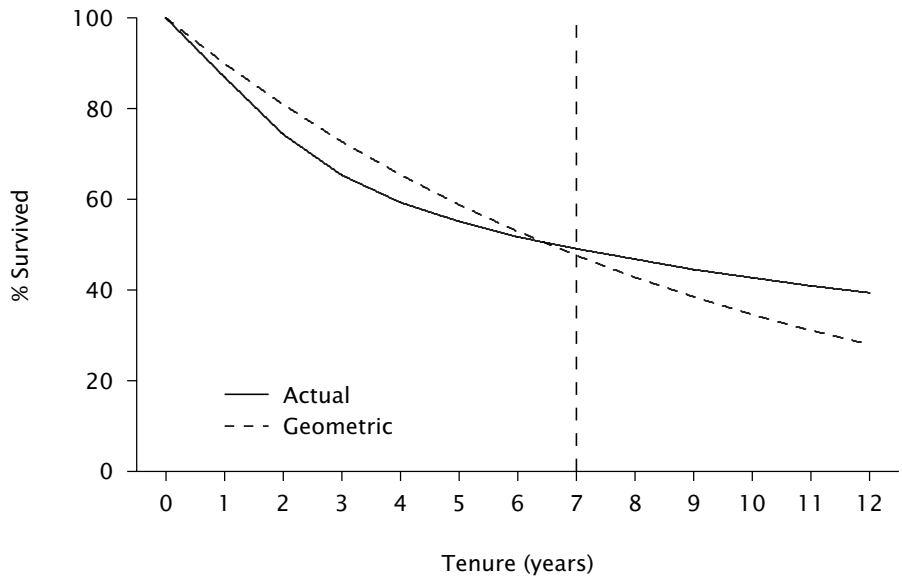
=$B$1*(1-$B$1)^(A8-1)

=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:

i. The duration of an individual customer's relationship with the firm is characterized by the (shifted) geometric distribution with parameter θ.

ii. 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 $B(\alpha, \beta)$ is defined by the integral

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

  and can be expressed in terms of gamma functions:

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

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

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

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

The Beta Distribution

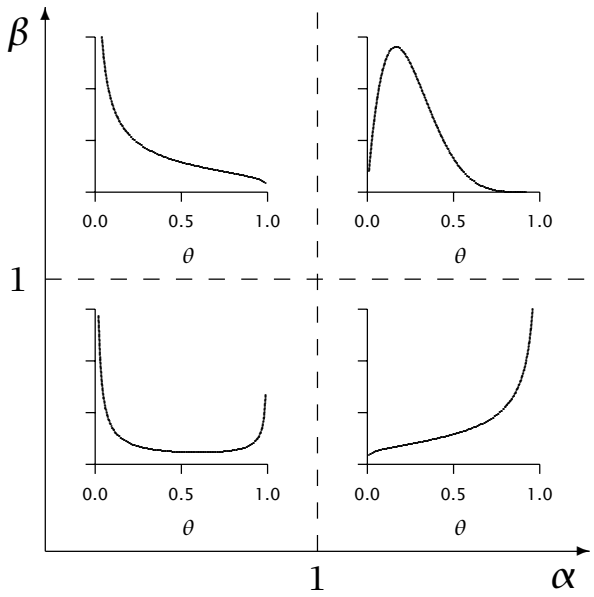
$$f(\theta \mid \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)}, \ 0 < \theta < 1.$$

- 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 (II)

For a randomly-chosen individual,

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

This is 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:

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

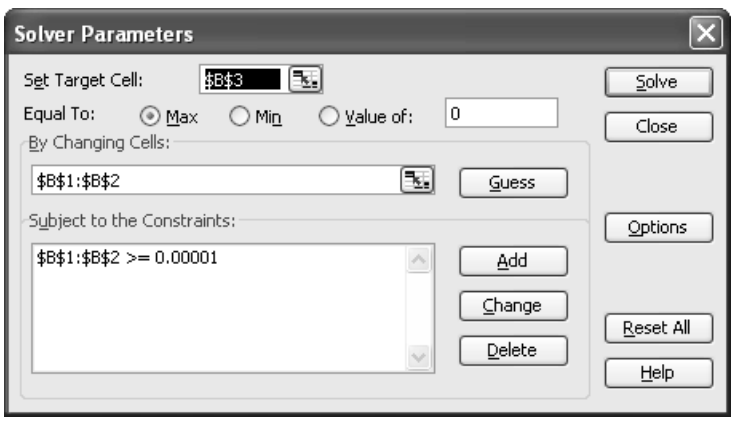
The maximum value of the log-likelihood function is $LL = -1611.16$, which occurs at $\hat{\alpha} = 0.668$ and $\hat{\beta} = 3.806$.

Estimating Model Parameters

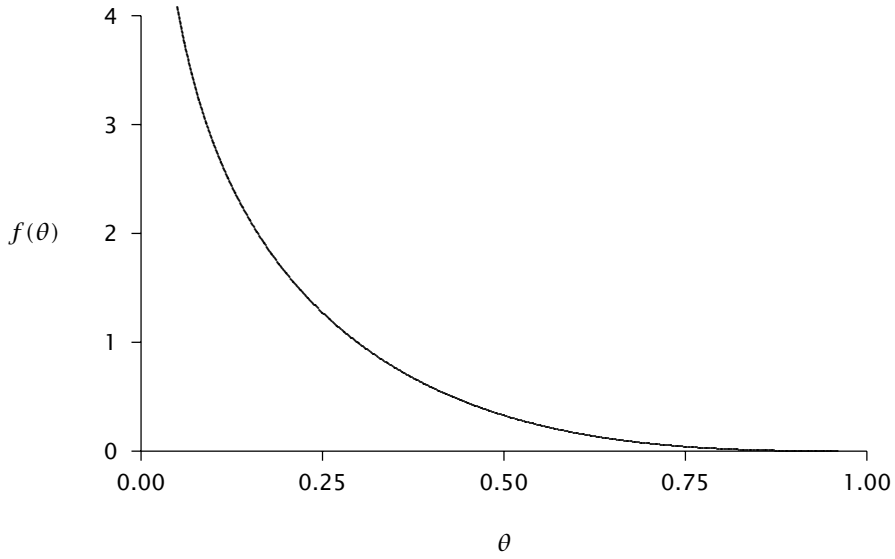
	A	B	C	D	E
1	alpha	1.000			
2	beta	1.000			
3	LL	-2115.55			
4					
5	Year	P(T=t)	# Cust.	# Lost	
6	0		1000		
7	1	0.5000	=B1/(B1+B2)	31	-90.8023
8	2	0.1667	743	126	-225.7617
9				90	-223.6416
10				60	-179.7439
11	5	0.0333	551	42	-142.8503
12	6	0.0238	517	34	-127.0808
13	7	0.0179	491	26	-104.6591
14					-1021.0058

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

Estimating Model Parameters

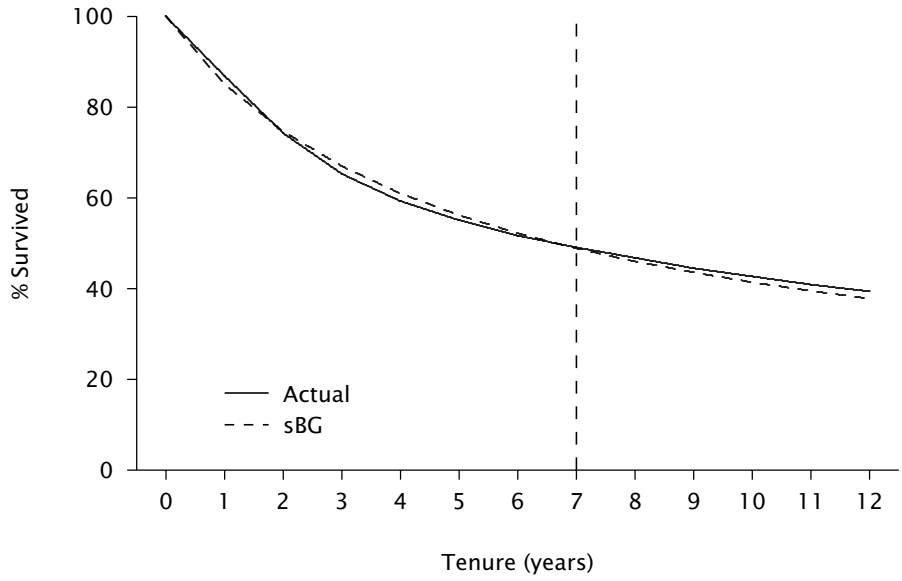


Estimated Distribution of Churn Probabilities



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

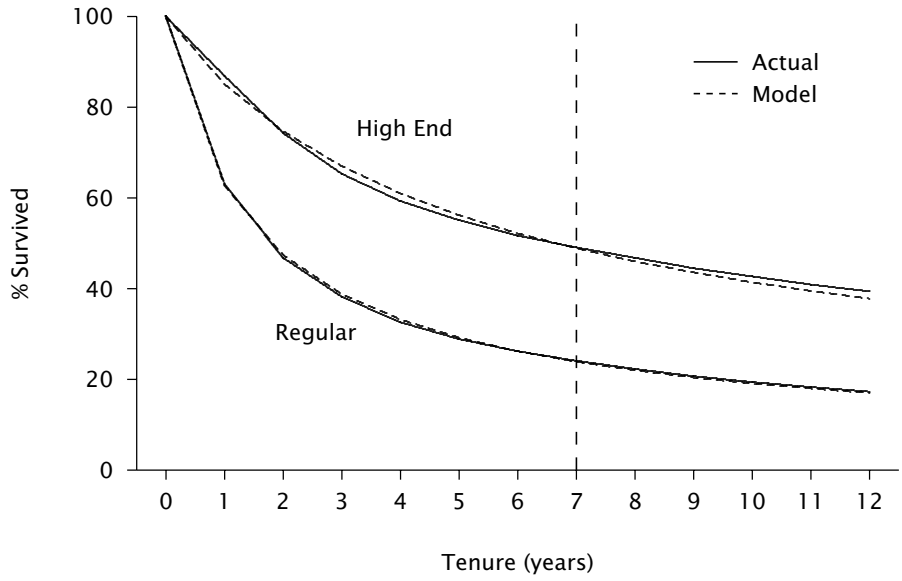
Survival Curve Projection



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



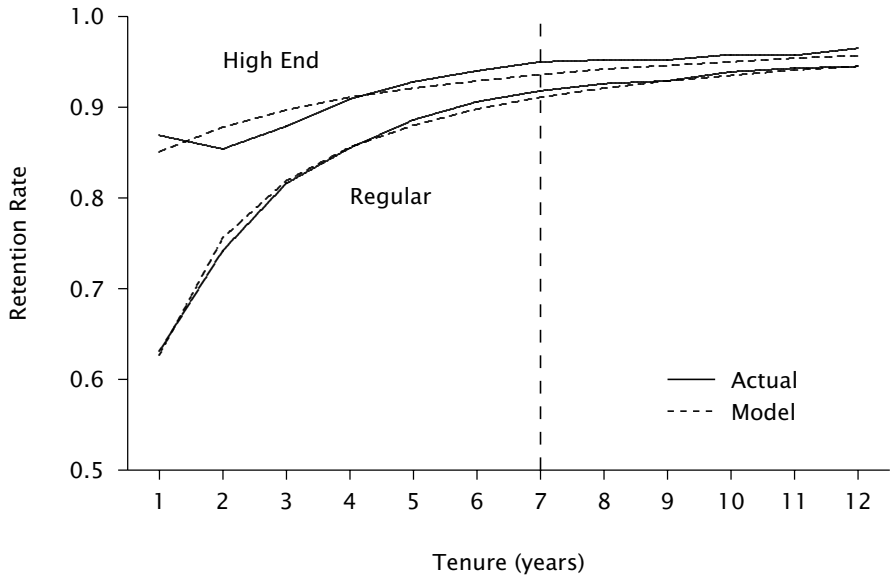
Implied Retention Rates

- 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:

$$r_t = \frac{P(T > t)}{P(T > t-1)} = \frac{\beta + t - 1}{\alpha + \beta + t - 1}.$$

- 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
- Modeling 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 (2005), "A Simple Probability Model for Projecting Customer Retention." [`http://brucehardie.com/papers/021/`]

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
  (Modeling Discrete-Time Duration Data)
- Problem 2: Predicting New Product Trial
  (Modeling Continuous-Time Duration Data)
- Problem 3: Estimating Billboard Exposures
  (Modeling Count Data)
- Problem 4: Test/Roll Decisions in Segmentation- based Direct Marketing
  (Modeling "Choice" Data)
- Problem 5: Characterizing the Purchasing of Hard-Candy
  (Introduction to Finite Mixture Models)
- Problem 6: Who is Visiting khakichinos.com?
  (Incorporating Covariates in Count Models)

Problem 2:
Predicting New Product Trial

(Modeling 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.

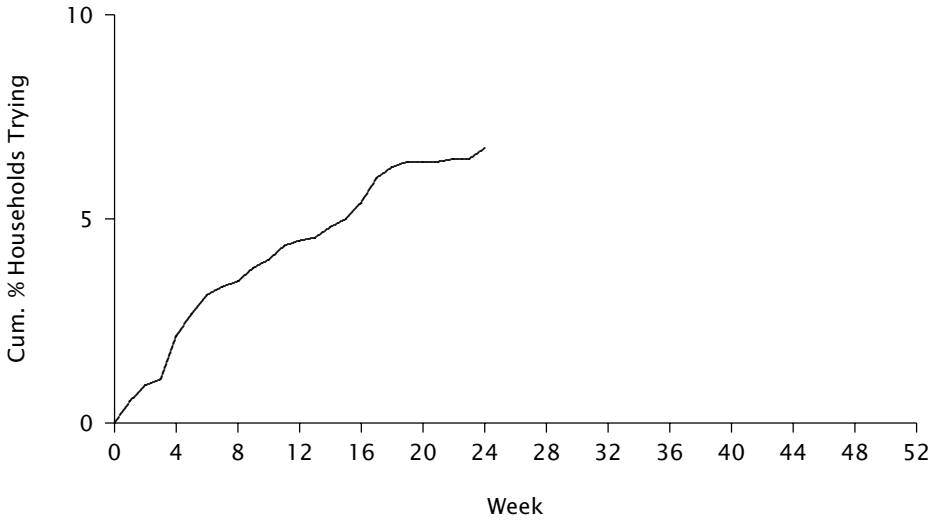
The product has now been under test for 24 weeks. On hand is a dataset documenting the number of households that have made a trial purchase by the end of each week. (The total size of the panel is 1499 households.)

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 ⇒ 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 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.

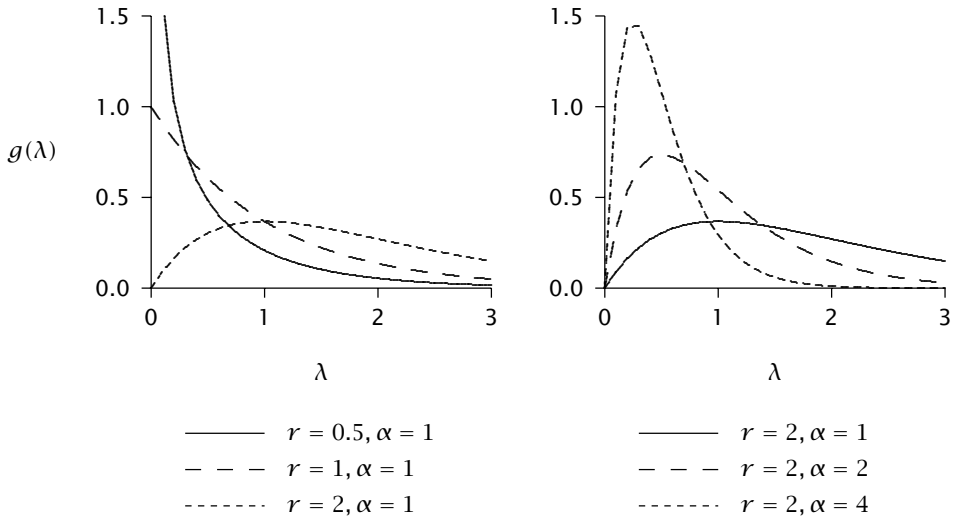
Distribution of Trial Rates

- 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.

Illustrative Gamma Density Functions



Market-Level Model

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

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

We call this the “exponential-gamma” model.

Estimating Model Parameters

The log-likelihood function is defined as:

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

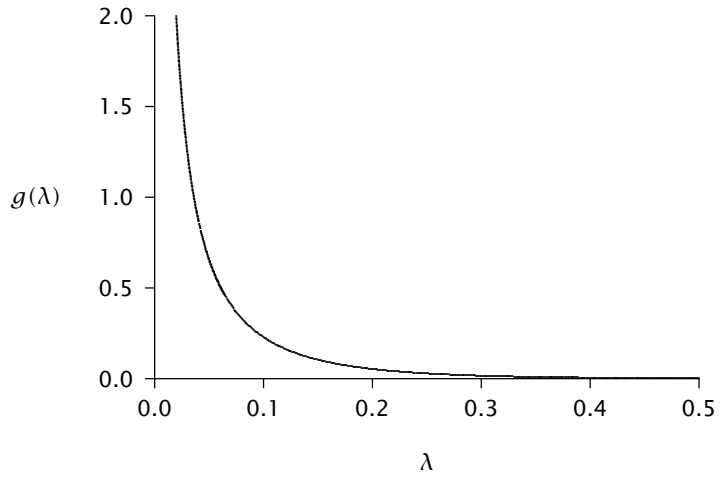
The maximum value of the log-likelihood function is $LL = -681.4$, which occurs at $\hat{r} = 0.050$ and $\hat{\alpha} = 7.973$.

Estimating Model Parameters

	A	B	C	D	E	F
1	Product:	Kiwi Bubbles			r	1.000
2	Panelists:	1499			alpha	1.000
3			=SUM(F6:F30)		LL	-4909.5
4		Cum_Trl				
5	Week	# HHs	Incr_Trl	P(T <= t)	P(try week t)	
6		=1-(F$2/(F$2+A6))^F$1	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	=C8*LN(E8)	-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				=(B2-B29)*LN(1-D29)		-4499.988

=D7-D6

Estimated Distribution of λ



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

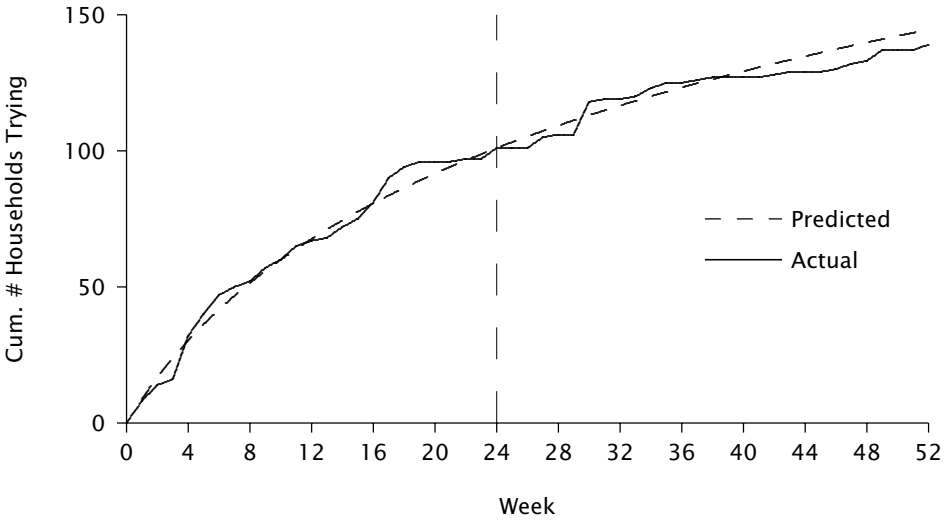
Forecasting Trial

- $F(t)$ 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

- Modeling 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

(Modeling 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

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.

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).

Model Development (I)

- 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:

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

- All individuals are assumed to have the same exposure rate.

Estimating Model Parameters

The log-likelihood function is defined as:

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

The maximum value of the log-likelihood function is $LL = -929.0$, which occurs at $\hat{\lambda} = 4.456$.

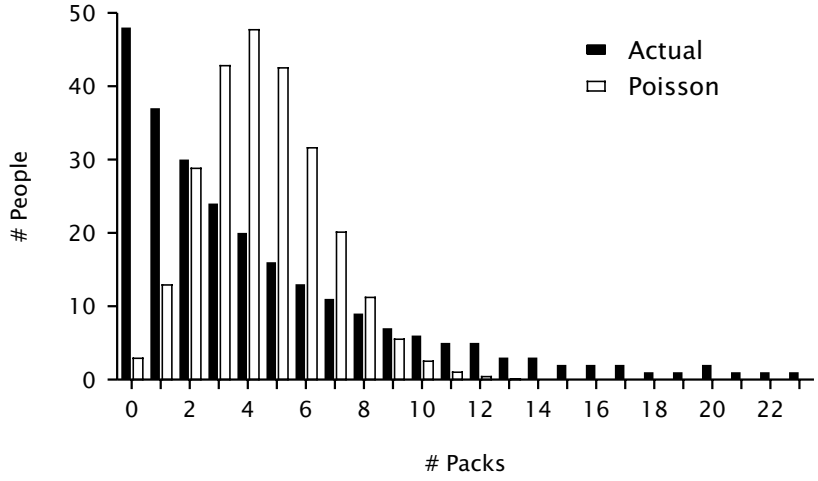
Estimating Model Parameters

	A	B	C	D
1	lambda	3.000		
2	LL	-1005.8	← =SUM(D5:D28)	
3				
4	x	f_x	P(X=x)	
5	0	48	0.04979	-144.00
6				-70.35
7				-44.88
8	3	24	0.22404	-35.90
9	4			-35.67
10	5	16	0.10082	-36.71
11	6	13	0.05041	-38.84
12	7	11	0.02160	-42.18
13	8	9	0.00810	-43.34
14	9	7	0.00270	-41.40
27	22	1	0.00000	-27.30
28	23	1	0.00000	-29.34

=POISSON(A5,$B$1,FALSE)

=B9*LN(C9)

Fit of the Poisson Model



Model Development (II)

- 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 (II)

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.

Mean of the NBD

We can derive an expression for the mean of the NBD *by conditioning*:

$$\begin{aligned} E(X) &= E\big[E(X \mid \lambda)\big] \\ &= \int_0^\infty E(X \mid \lambda) g(\lambda \mid r, \alpha)\, d\lambda \\ &= \frac{r}{\alpha}. \end{aligned}$$

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:

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

Estimating Model Parameters

The log-likelihood function is defined as:

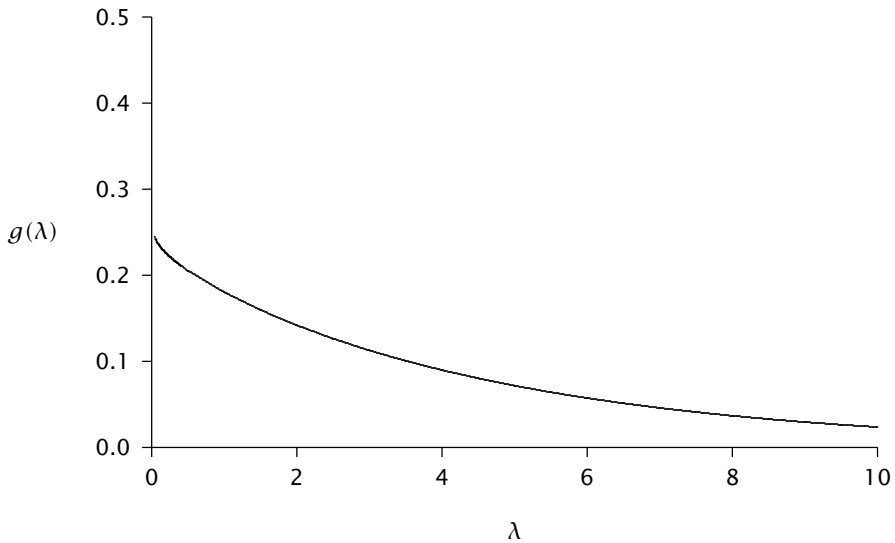
$$\begin{aligned} LL(r,\alpha|\text{data}) = {} & 48 \times \ln[P(X=0)] \; + \\ & 37 \times \ln[P(X=1)] \; + \\ & 30 \times \ln[P(X=2)] \; + \\ & \ldots \; + \\ & 1 \times \ln[P(X=23)] \end{aligned}$$

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	=(B2/(B2+1))^B1	
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		=C6*($B$1+A7-1)/(A7*($B$2+1))		-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

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

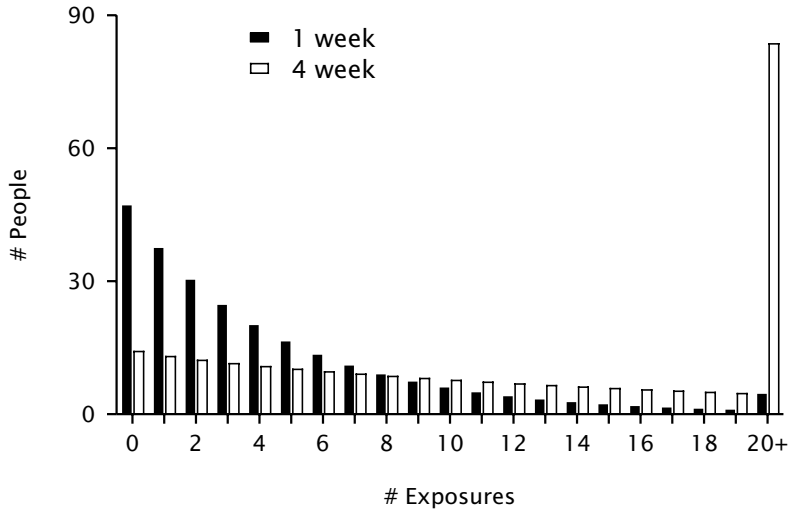
- The distribution of exposures at the population-level is given by:

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

- 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:
  - Reach $= 1 - P(X(t) = 0)$
    $= 94.4\%$
  - Frequency $= E[X(t)]/(1 - P(X(t) = 0))$
    $= 18.9$
  - GRPs $= 100 \times E[X(t)]$
    $= 1782$

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

(Modeling "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.

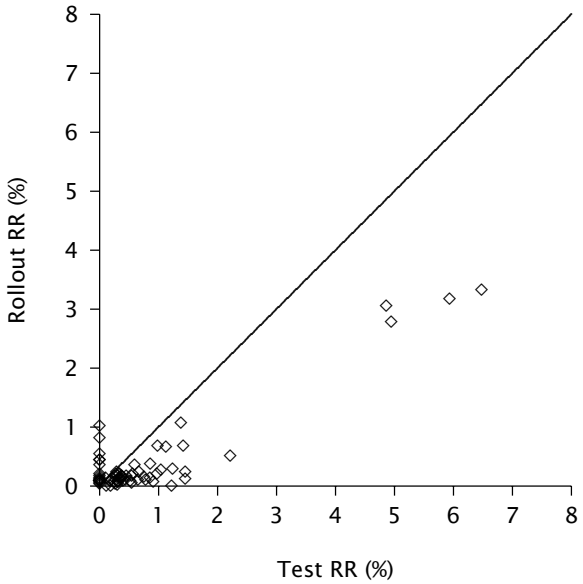
Standard approach:

- Rollout to all segments with

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

- 51 segments pass this hurdle

Test vs. Actual Response Rate



Modeling 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:

$$LL(\alpha,\beta|\text{data}) = \sum_{s=1}^{126} \ln[P(X_s = x_s|m_s,\alpha,\beta)]$$

$$= \sum_{s=1}^{126} \ln\left[\frac{m_s!}{(m_s-x_s)!\,x_s!}\underbrace{\frac{\Gamma(\alpha+x_s)\Gamma(\beta+m_s-x_s)}{\Gamma(\alpha+\beta+m_s)}}_{B(\alpha+x_s,\beta+m_s-x_s)}\underbrace{\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}}_{1/B(\alpha,\beta)}\right]$$

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

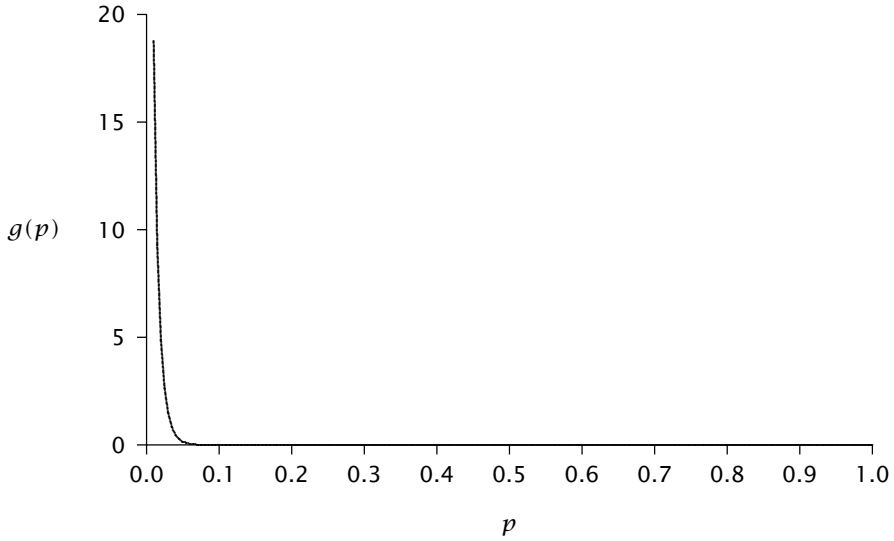
Estimating Model Parameters

	A	B	C	D	E
1	alpha	1.000		B(alpha,beta)	1.000
2	beta	1.000			
3	LL	-718.9	← =SUM(E6:E131)		
4					
5	Segment	m_s	x_s	P(X=x|m)	
6	1	34	0	0.02857	-3.555
7					-4.635
8					-3.989
9					-4.984
10					-7.135
11	6	144	7	0.00690	-4.977
12	7	1235	80		-7.120
13	8	573	34		-6.353
14	9	1083	24	0.00092	-6.988
130	125	383	0	0.00260	-5.951
131	126	404	0	0.00247	-6.004

=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

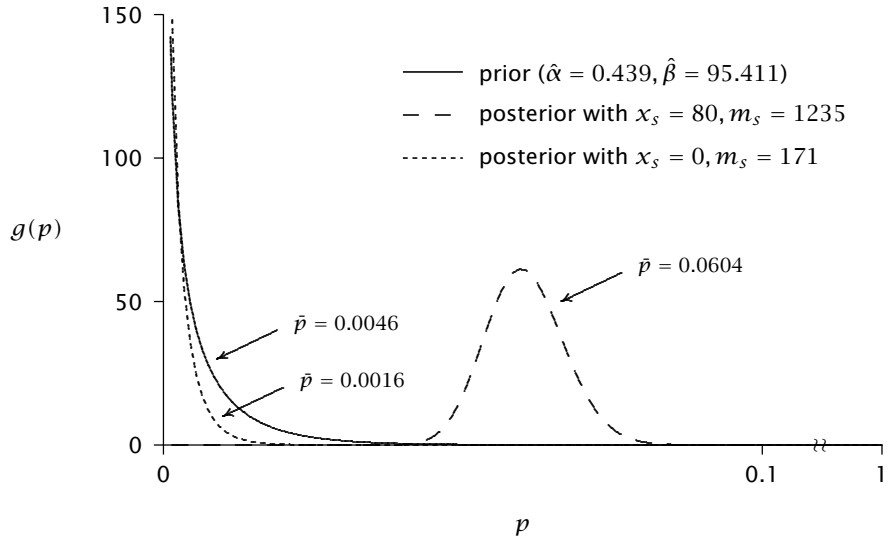
For the beta-binomial model, we have:

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

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

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:

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

- 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.)

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.

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.

Problem 5:
Characterizing the Purchasing of Hard-Candy

(Introduction to Finite Mixture Models)

Distribution of Hard-Candy Purchases

# Packs	# People	# Packs	# People
0	102	11	10
1	54	12	10
2	49	13	3
3	62	14	3
4	44	15	5
5	25	16	5
6	26	17	4
7	15	18	1
8	15	19	2
9	10	20	1
10	10		

Source: Dillon and Kumar (1994)

Fitting the NBD

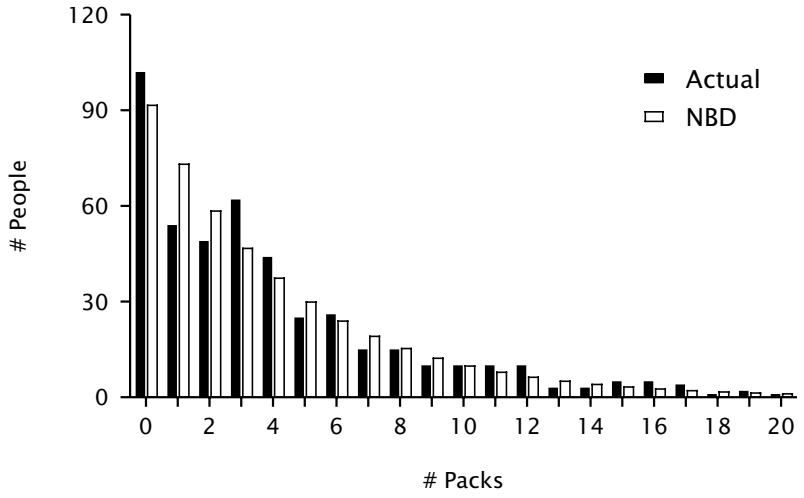
	A	B	C	D	E	F	G	H	I	J
1	r	0.998								
2	alpha	0.250								
3	LL	-1140.02								
4										
5	# Packs	Observed	P(X=x)	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
6	0	102	0.20073	-163.79	91.5		0	102	91.5	1.20
7	1	54	0.16021	-98.89	73.1		1	54	73.1	4.97
8	2	49	0.12802	-100.72	58.4		2	49	46.7	1.51
9	3	62	0.10234	-141.32	46.7		3	62	37.3	5.04
10	4	44	0.08183	-110.14	37.3		4	44	29.8	1.20
11	5	25	0.06543	-68.17	29.8		5	25	23.9	0.78
12	6	26	0.05233	-76.71	23.9		6	26	19.1	0.19
13	7	15	0.04185	-47.60	19.1		7	15	15.3	0.87
14	8	15	0.03347	-50.96	15.3		8	15	12.2	0.00
15	9	10	0.02677	-36.20	12.2		9	10	9.8	0.40
16	10	10	0.02141	-38.44	9.8		10	10	7.8	0.01
17	11	10	0.01713	-40.67	7.8		11	10	6.2	0.61
18	12	10	0.01370	-42.90	6.2		12	10	5.0	2.25
19	13	3	0.01096	-13.54	5.0		13	3	4.0	0.80
20	14	3	0.00876	-14.21	4.0		14	3	11.8	0.25
21	15	5	0.00701	-24.80	3.2		15+	18		3.27
22	16	5	0.00561	-25.92	2.6				# params	23.35
23	17	4	0.00449	-21.63	2.0				df	2
24	18	1	0.00359	-5.63	1.6				p-value	13
25	19	2	0.00287	-11.71	1.3					0.038
26	20	1	0.00230	-6.08	1.0					
27		456								

=B$27*C6

=(H6-I6)^2/I6

=CHIDIST(J22,J25)

Fit of the NBD



The Zero-Inflated NBD Model

Because of the "excessive" number of zeros, let us consider the zero-inflated NBD (ZNBD) model:

- a proportion π of the population never buy hard-candy
- the visiting behavior of the "ever buyers" can be characterized by the NBD model

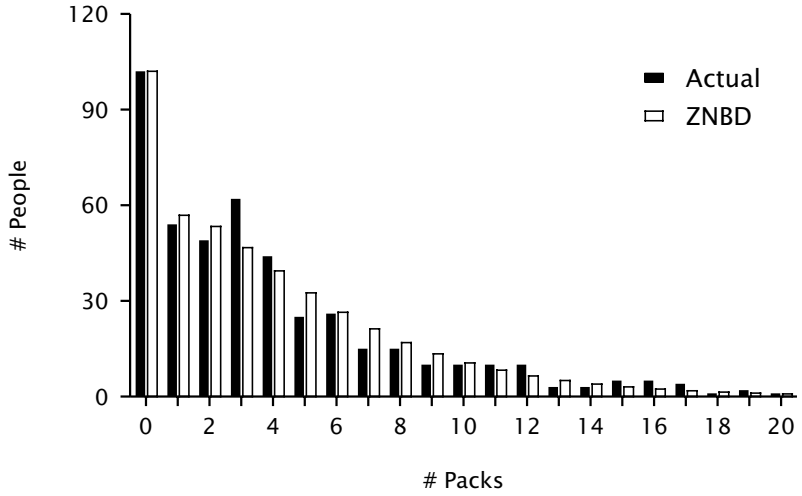
$$P(X = x) = \delta_{x=0}\pi + (1-\pi) \times \frac{\Gamma(r+x)}{\Gamma(r)x!}\left(\frac{\alpha}{\alpha+1}\right)^r\left(\frac{1}{\alpha+1}\right)^x$$

This is sometimes called the "NBD with hard-core non-buyers" model.

Fitting the ZNBD

	A	B	C	D	E	F	G	H	I	J	K
1	r	1.504									
2	alpha	0.334									
3	pi	0.113									
4	LL	-1136.17									
5											
6			P(X=x)								
7	# Packs	Observed	NBD	ZNBD	LL	Expected		# Packs	Observed	Expected	(O-E)^2/E
8	0	102	0.12468	0.22368	-152.75	102.0		0	102	102.0	0.00
9	1	54	0.14054	0.12465	-112.44	56.8		1	54	56.8	0.14
10	2	49	0.13188	0.11	=(A8=0)*B$3+(1-B$3)*C8			2	49	53.3	0.35
11	3	62	0.11545	0.10239	-141.29	46.7		3	62	46.7	5.02
12	4	44	0.09743	0.08641	-107.74	39.4		4	44	39.4	0.54
13	5	25	0.08039	0.07130	-66.02	32.5		5	25	32.5	1.74
14	6	26	0.06531	0.05793	-74.06	26.4		6	26	26.4	0.01
15	7	15	0.05248	0.04654	-46.01	21.2		7	15	21.2	1.82
16	8	15	0.04181	0.03708	-49.42	16.9		8	15	16.9	0.22
17	9	10	0.03309	0.02935	-35.28	13.4		9	10	13.4	0.86
18	10	10	0.02605	0.02311	-37.68	10.5		10	10	10.5	0.03
19	11	10	0.02042	0.01811	-40.11	8.3		11	10	8.3	0.37
20	12	10	0.01595	0.01415	-42.58	6.5		12	10	6.5	1.95
21	13	3	0.01242	0.01101	-13.53	5.0		13	3	5.0	0.81
22	14	3	0.00964	0.00855	-14.28	3.9		14	3	3.9	0.21
23	15	5	0.00747	0.00663	-25.08	3.0		15+	18	10.4	5.48
24	16	5	0.00578	0.00512	-26.37	2.3					19.54
25	17	4	0.00446	0.00395	-22.13	1.8					
26	18	1	0.00343	0.00305	-5.79	1.4				# params	3
27	19	2	0.00264	0.00234	-12.11	1.1				df	12
28	20	1	0.00203	0.00180	-6.32	0.8					
29		456								p-value	0.076

Fit of the ZNBD



What is Wrong With the NBD Model?

The assumptions underlying the model could be wrong on two accounts:

i. at the individual-level, the number of purchases is not Poisson distributed

ii. purchase rates (λ) are not gamma-distributed

Relaxing the Gamma Assumption

- Replace the continuous distribution with a discrete distribution by allowing for multiple (discrete) segments each with a different (latent) buying rate:

$$P(X = x) = \sum_{s=1}^{S} \pi_s P(X = x|\lambda_s), \quad \sum_{s=1}^{S} \pi_s = 1$$

- This is called a finite mixture model.
- We often reparameterize the mixing proportions for computational convenience:

$$\pi_s = \frac{\exp(\theta_s)}{\sum_{s'=1}^{S} \exp(\theta_{s'})}, \quad \theta_S = 0.$$

Fitting the One-Segment Model

	A	B	C	D
1	lambda	3.991		
2	LL	-1545.00		
3				
4	# Packs	Observed	P(X=x)	LL
5	0	102	0.01848	-407.11
6	1	54	0.07375	-140.78
7	2	49	0.14717	-93.89
8	3	62	0.19579	-101.10
9	4	44	0.19536	-71.85
10	5	25	0.15595	-46.46
25	20	1	0.00000	-18.64
26		456		

Fitting the Two-Segment Model

	A	B	C	D	E	F
1	lambda_1	1.802				
2	lambda_2	9.121				
3	pi	0.701				
4	LL	-1188.83				
5						
6	# Packs	Observed	Seg1	Seg2	P(X=x)	LL
7	0	102	0.16494	0.00011	0.11564	-220.04
8			725	0.00100	0.20864	-84.63
9				0.00455	0.189	
10						-133.07
11	4	44	0.07249	0.0		23.61
12	5	25	0.02613	0.05753	0.03552	-83.44
27	20	1	0.00000	0.00071	0.00021	-8.45
28		456				

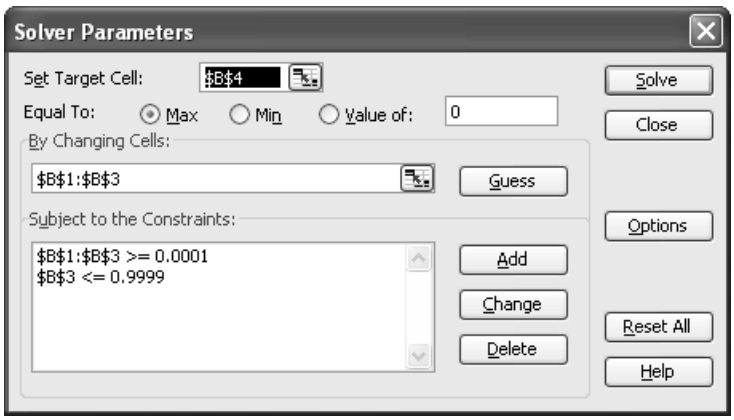
=POISSON(A7,B$1,FALSE)

=POISSON(A7,B$2,FALSE)

=B$3*C7+(1-B$3)*D7

=B7*LN(E7)

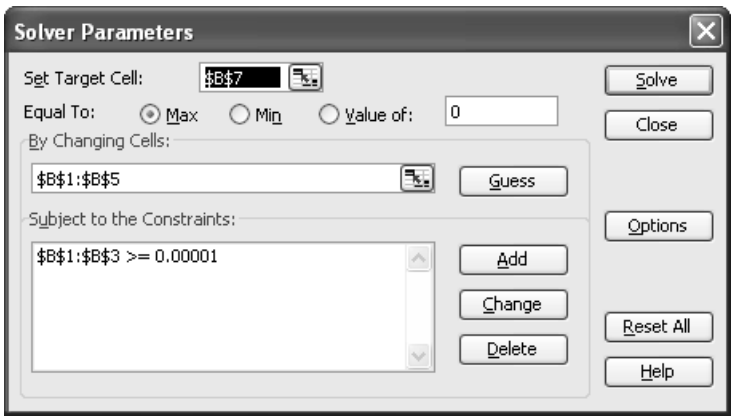
Fitting the Two-Segment Model



Fitting the Three-Segment Model

	A	B	C	D	E	F	G
1	lambda_1	3.483					
2	lambda_2	11.216					
3	lambda_3	0.291					
4	theta_1	0.674	1.963	← =EXP(B4)			
5	theta_2	-0.430	0.650				
6	theta_3	0	1.000				
7	LL	-1132.04		=C4/SUM(C4:C6)			
8				↙			
9			0.543	0.180	0.277		
10	# Packs	Observed	Seg1	Seg2	Seg3	P(X=x)	LL
11	0	102	0.03071	0.00001	0.74786	0.22367	-152.76
12	1	54	0.10696	0.00015	0.21728	0.11827	-115.28
13	2	49	=SUMPRODUCT(C$9:E$9,C11:E11)			1009	-108.12
14	3	62	0.21629	0.00317	0.00306	0.11892	-132.02
15	4	44	0.18835	0.00887	0.00022	0.10399	-99.59
16	5	25	0.13122	0.01991	0.00001	0.07487	-64.80
31	20	1	0.00000	0.00549	0.00000	0.00099	-6.92
32		456					

Fitting the Three-Segment Model



Fitting the Four-Segment Model

	A	B	C	D	E	F	G	H
1	lambda_1	3.002						
2	lambda_2	0.205						
3	lambda_3	7.418						
4	lambda_4	12.873						
5	theta_1	1.598	4.943					
6	theta_2	0.876	2.401					
7	theta_3	0.398	1.489					
8	theta_4	0	1.000					
9	LL	-1130.07						
10								
11			0.503	0.244	0.151	0.102		
12	# Packs	Observed	Seg1	Seg2	Seg3	Seg4	P(X=x)	LL
13	0	102	0.04969	0.81487	0.00060	0.00000	0.22406	-152.58
14	1	54	0.14917	0.16683	0.00445	0.00003	0.11641	-116.14
15	2	49	0.22390	0.01708	0.01652	0.00021	0.11925	-104.20
16	3	62	0.22404	0.00117	0.04084	0.00091	0.11919	-131.88
17	4	44	0.16814	0.00006	0.07574	0.00294	0.09631	-102.97
18	5	25	0.10095	0.00000	0.11237	0.00756	0.06853	-67.01
33	20	1	0.00000	0.00000	0.00006	0.01647	0.00168	-6.39
34		456						

Parameter Estimates

	Seg 1	Seg 2	Seg 3	Seg 4	LL
λ	3.991				-1545.00
λ_s	1.802	9.121			-1188.83
π_s	0.701	0.299			
λ_s	0.291	3.483	11.216		-1132.04
π_s	0.277	0.543	0.180		
λ_s	0.205	3.002	7.418	12.873	-1130.07
π_s	0.244	0.503	0.151	0.102	

How Many Segments?

- Controlling for the extra parameters, is an $S+1$ segment model better than an S segment model?
- We can't use the likelihood ratio test because its properties are violated
- It is standard practice to use "information-theoretic" model selection criteria
- A common measure is the Bayesian information criterion:

  $$\text{BIC} = -2LL + p\ln(N)$$

  where p is the number of parameters and N is the sample size
- Rule: choose S to minimize BIC

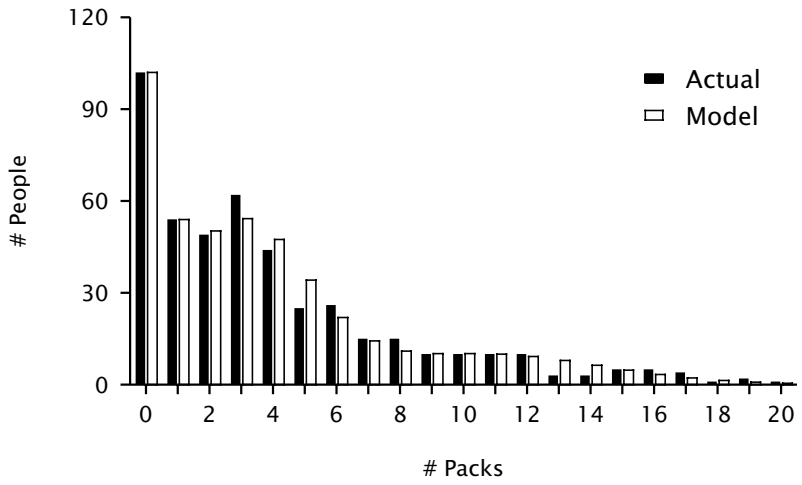
Summary of Model Fit

Model	LL	# params	BIC	χ^2 p-value
NBD	−1140.02	2	2292.29	0.04
ZNBD	−1136.17	3	2290.70	0.08
Poisson	−1545.00	1	3096.12	0.00
2 seg Poisson	−1188.83	3	2396.03	0.00
3 seg Poisson	−1132.04	5	2294.70	0.22
4 seg Poisson	−1130.07	7	2303.00	0.33

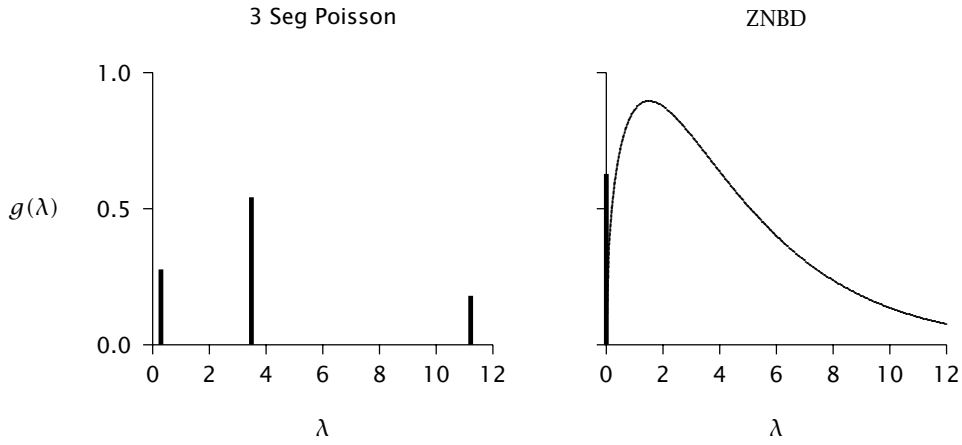
LatentGOLD Results

	Seg 1	Seg 2	Seg 3	Seg 4	*LL*
mean	3.991				−1545.00
class size	1.000				
mean	1.801	9.115			−1188.83
class size	0.700	0.300			
mean	3.483	0.291	11.210		−1132.04
class size	0.542	0.277	0.181		
mean	2.976	0.202	7.247	12.787	−1130.07
class size	0.500	0.243	0.156	0.106	

Fit of the Three-Segment Poisson Model



Implied Heterogeneity Distribution



$\hat{\pi} = 0.113, \hat{r} = 1.504, \hat{\alpha} = 0.334$

Classification Using Bayes Theorem

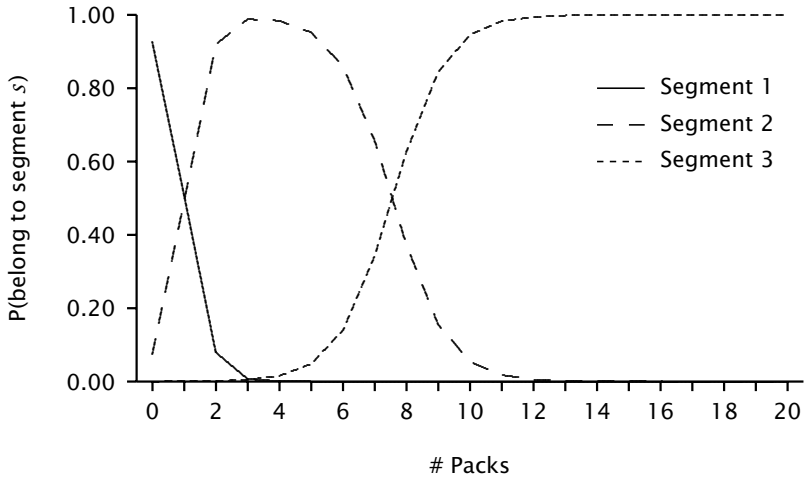
To which "segment" of the mixing distribution does each observation x belong?

- π_s can be interpreted as the prior probability of λ_s
- By Bayes theorem,

$$P(s \mid X = x) = \frac{P(X = x \mid \lambda_s)\pi_s}{\sum_{s'=1}^{S} P(X = x \mid \lambda_{s'})\pi_{s'}},$$

  which can be interpreted as the posterior probability of λ_s

Posterior Probabilities



Conditional Expectations

What is the expected purchase quantity over the next month for a customer who purchased seven packs last week?

$$
\begin{aligned}
E[X(4)] &= E[X(4)|\text{seg } 1]\,P(\text{seg } 1|X=7) \\
&\quad + E[X(4)|\text{seg } 2]\,P(\text{seg } 2|X=7) \\
&\quad + E[X(4)|\text{seg } 3]\,P(\text{seg } 3|X=7) \\
&= (4 \times 0.291) \times 0.0000 \\
&\quad + (4 \times 3.483) \times 0.6575 \\
&\quad + (4 \times 11.216) \times 0.3425 \\
&= 24.5
\end{aligned}
$$

… or 13.9 with “hard assignment” to segment 2.

Concepts and Tools Introduced

- Finite mixture models
- Discrete vs. continuous mixing distributions
- Probability models for classification

Further Reading

Dillon, Wiliam R. and Ajith Kumar (1994), "Latent Structure and Other Mixture Models in Marketing: An Integrative Survey and Overview," in Richard P. Bagozzi (ed.), *Advanced Methods of Marketing Research*, Oxford: Blackwell.

McLachlan, Geoffrey and David Peel (2000), *Finite Mixture Models*, New York: John Wiley & Sons.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Problem 6:
Who is Visiting khakichinos.com?

(Incorporating Covariates in Count Models)

Background

Khaki Chinos, Inc. is an established clothing catalog company with an online presence at khakichinos.com. While the company is able to track the online *purchasing* behavior of its customers, it has no real idea as to the pattern of *visiting* behaviors by the broader Internet population.

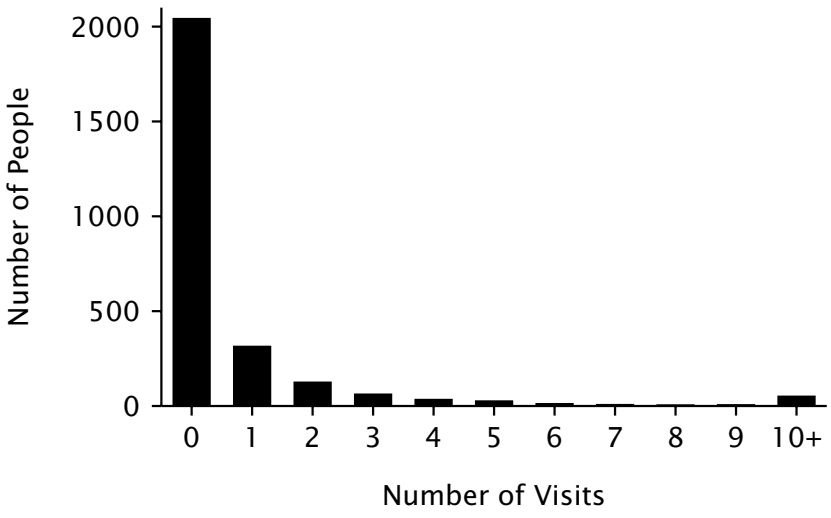
In order to gain an understanding of the aggregate visiting patterns, some Media Metrix panel data has been purchased. For a sample of 2728 people who visited an online apparel site at least once during the second-half of 2000, the dataset reports how many visits each person made to the khakichinos.com web site, along with some demographic information.

Management would like to know whether frequency of visiting the web site is related to demographic characteristics.

Raw Data

ID	# Visits	ln(Income)	Sex	ln(Age)	HH Size
1	0	11.38	1	3.87	2
2	5	9.77	1	4.04	1
3	0	11.08	0	3.33	2
4	0	10.92	1	3.95	3
5	0	10.92	1	2.83	3
6	0	10.92	0	2.94	3
7	0	11.19	0	3.66	2
8	1	11.74	0	4.08	2
9	0	10.02	0	4.25	1
	...				

Distribution of Visits



Modeling Count Data

Recall the NBD:

- At the individual-level, $Y \sim \text{Poisson}(\lambda)$
- λ is distributed across the population according to a gamma distribution with parameters r and α

$$P(Y = y) = \frac{\Gamma(r + y)}{\Gamma(r)y!}\left(\frac{\alpha}{\alpha + 1}\right)^r \left(\frac{1}{\alpha + 1}\right)^y$$

Observed vs. Unobserved Heterogeneity

Unobserved Heterogeneity:

- People differ in their mean (visiting) rate λ
- To account for heterogeneity in λ, we assume it is distributed across the population according to some (parametric) distribution
- But there is no attempt to *explain* how people differ in their mean rates

Observed Heterogeneity:

- We observe how people differ on a set of observable independent (explanatory) variables
- We explicitly link an individual's λ to her observable characteristics

The Poisson Regression Model

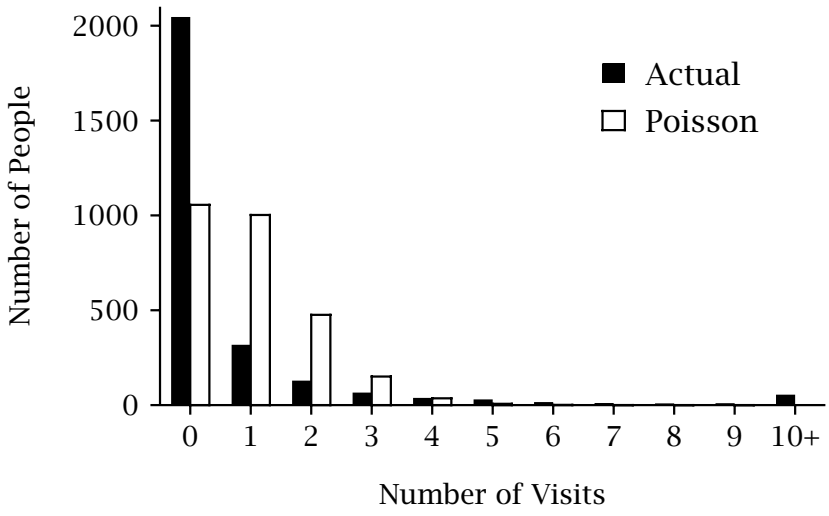
- Let the random variable Y_i denote the number of times individual i visits the site in a unit time period
- At the individual-level, Y_i is assumed to be distributed Poisson with mean λ_i:

$$P(Y_i = y|\lambda_i) = \frac{\lambda_i^y e^{-\lambda_i}}{y!}$$

- An individual's mean is related to her observable characteristics through the function

$$\lambda_i = \lambda_0 \exp(\boldsymbol{\beta}' \mathbf{x}_i)$$

Fit of the Poisson Model



$$\hat{\lambda} = 0.949, LL = -6378.6$$

Fitting the Poisson Regression Model

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	0.0439			LL	-6291.497				
2	B_inc	0.0938		{=TRANSPOSE(B2:B5)}						
3	B_sex	0.0043								
4	B_age	0.5882								
5	B_size	-0.0359								
6				0.0938	0.0043	0.5882	-0.0359			
7										
8	ID	Total		Income	Sex	Age	HH Size	lambda	P(Y=y)	ln[P(Y=y)]
9	1	0		11.38	1	3.87	2	1.16317	0.31249	-1.163
10	2						1	1.14695	0.00525	-5.249
11	3			=B$1*EXP(SUMPRODUCT(D$6:G$6,D9:G9))			2	0.82031	0.44030	20
12	4	0		10.92	1	=H9^B9*EXP(-H9)/FACT(B9)			0.32430	-1.126
13	5	0		10.92	1	2.83	3	0.58338	0.55801	-0.583
14	6	0		10.92	0	2.94	3	0.62017	0.53785	-0.620
15	7	0		11.19	0	3.66	2	1.00712	0.36527	-1.007
16	8	1		11.74	0	4.08	2	1.35220	0.34977	-1.050
17	9	0		10.02	0	4.25	1	1.31954	0.26726	-1.320
18	10	0		10.92	0	3.85	3	1.05656	0.34765	-1.057
2735	2727	0		10.53	1	2.89	4	0.56150	0.57035	-0.561
2736	2728	0		11.74	1	2.83	3	0.63010	0.53254	-0.630

=LN(I9)

Poisson Regression Results

Variable	Coefficient
λ_0	0.0439
Income	0.0938
Sex	0.0043
Age	0.5882
HH Size	−0.0359
LL	−6291.5
LL_{Poiss}	−6378.6
LR (df = 4)	174.2

Comparing Expected Visit Behavior

	Person A	Person B
Income	59,874	98,716
Sex	M	F
Age	55	33
HH Size	4	2

Who is less likely to have visited the web site?

$$\begin{aligned}\lambda_A &= 0.0439 \times \exp\left(0.0938 \times \ln(59,874) + 0.0043 \times 0\right.\\ &\quad \left. + 0.5882 \times \ln(55) - 0.0359 \times 4\right)\\ &= 1.127\\ \lambda_B &= 0.0439 \times \exp\left(0.0938 \times \ln(98,716) + 0.0043 \times 1\right.\\ &\quad \left. + 0.5882 \times \ln(33) - 0.0359 \times 2\right)\\ &= 0.944\end{aligned}$$

Is $\boldsymbol{\beta}$ Different from 0?

Consider two models, A and B:

> If we can arrive at model B by placing k constraints on the parameters of model A, we say that model B is *nested* within model A.

The Poisson model is nested within the Poisson regression model by imposing the constraint $\boldsymbol{\beta} = \mathbf{0}$.

We use the *likelihood ratio test* to determine whether model A, which has more parameters, fits the data better than model B.

The Likelihood Ratio Test

- The null hypothesis is that model A is not different from model B
- Compute the test statistic

$$LR = -2(LL_B - LL_A)$$

- Reject null hypothesis if $LR > \chi^2_{.05,k}$

Computing Standard Errors

- Excel
  - indirectly via a series of likelihood ratio tests
  - easily computed from the Hessian matrix (computed using difference approximations)
- General modeling environments (e.g., MATLAB, Gauss)
  - easily computed from the Hessian matrix (as a by-product of optimization or computed using difference approximations)
- Advanced statistics packages (e.g., Limdep, R, S-Plus)
  - they come for free

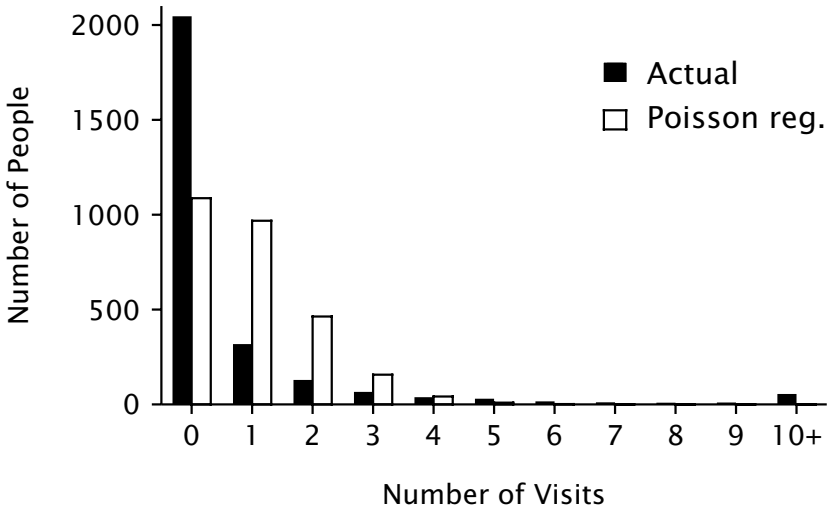
S-Plus Poisson Regression Results

```
Coefficients:
                  Value Std. Error     t value
(Intercept) -3.126238804 0.40578080  -7.7042552
     Income  0.093828021 0.03436347   2.7304580
        Sex  0.004259338 0.04089411   0.1041553
        Age  0.588249213 0.05472896  10.7484079
    HH Size -0.035907406 0.01528397  -2.3493511
```


Limdep Poisson Regression Results

```
+---------+--------------+----------------+--------+
|Variable | Coefficient  | Standard Error |b/St.Er.|
+---------+--------------+----------------+--------+
 Constant -3.122103284      .40565119       -7.697
 INCOME    .9305546493E-01  .34332533E-01    2.710
 SEX       .4312514407E-02  .40904265E-01     .105
 AGE       .5893014445      .54790230E-01   10.756
 HH SIZE  -.3577795361E-01  .15287122E-01   -2.340
```

Fit of the Poisson Regression



The ZIP Regression Model

Because of the "excessive" number of zeros, let us consider the zero-inflated Poisson (ZIP) regression model:

- a proportion π of those people who go to online apparel sites will never visit khakichinos.com
- the visiting behavior of the "ever visitors" can be characterized by the Poisson regression model

$$P(Y_i = y) = \delta_{y=0}\pi + (1-\pi) \times \frac{[\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)]^y e^{-\lambda_0 \exp(\boldsymbol{\beta}'\mathbf{x}_i)}}{y!}$$

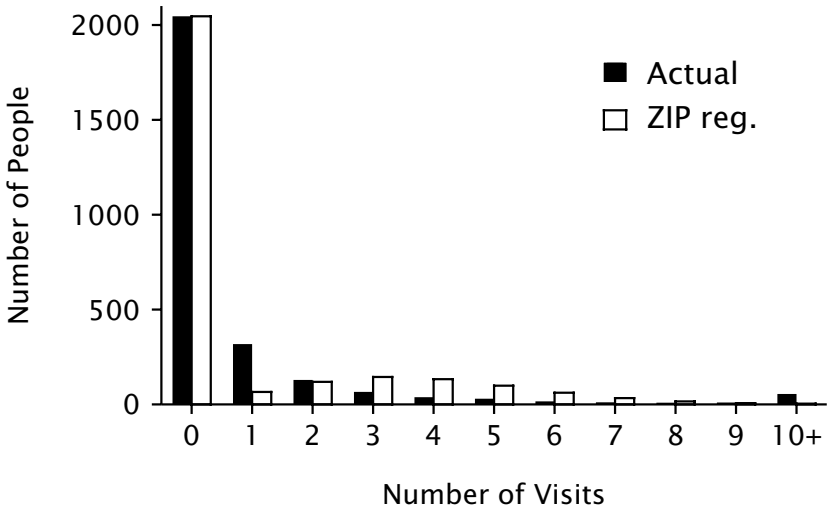
Fitting the ZIP Regression Model

	A	B	C	D	E	F	G	H	I	J
1	\lambda_0	6.6231			LL	-4297.472				
2	pi	0.7433								
3	B_inc	-0.0891								
4	B_sex	-0.1327								
5	B_age	0.1141								
6	B_size	0.0196								
7				-0.0891	-0.1327	0.1141	0.0196			
8										
9	ID	Total		Income	Sex	Age	HH Size	lambda	P(Y=y)	ln[P(Y=y)]
10	1	0		11.38	1	3.87	2	3.40193	0.75184	-0.285
11	2	5		9.77	1	4.04	1	3.92698	0.03936	-3.235
12	3	0		=IF(B10=0,B$2,0)+(1-B$2)*H10^B10*EXP(-H10)/FACT(B10)					2	-0.289
13	4	0		10.92	1	3.95	3	3.64889	0.74996	-0.288
14	5	0		10.92	1	2.83	3	3.21182	0.75363	-0.283
15	6	0		10.92	0	2.94	3	3.71435	0.74954	-0.288
16	7	0		11.19	0	3.66	2	3.85775	0.74871	-0.289
17	8	1		11.74	0	4.08	2	3.85266	0.02099	-3.864
18	9	0		10.02	0	4.25	1	4.48880	0.74617	-0.293
19	10	0		10.92	0	3.85	3	4.11879	0.74746	-0.291
2736	2727	0		10.53	1	2.89	4	3.41119	0.75176	-0.285
2737	2728	0		11.74	1	2.83	3	2.98515	0.75626	-0.279

ZIP Regression Results

Variable	Coefficient
λ_0	6.6231
Income	−0.0891
Sex	−0.1327
Age	0.1141
HH Size	0.0196
π	0.7433
LL	−4297.5
$LL_{\text{Poiss reg}}$	−6291.5
LR (df = 1)	3988.0

Fit of the ZIP Regression



NBD Regression

The explanatory variables may not fully capture the differences among individuals

To capture the remaining (unobserved) component of differences among individuals, let λ_0 vary across the population according to a gamma distribution with parameters r and α:

$$P(Y_i = y) = \frac{\Gamma(r+y)}{\Gamma(r)y!}\left(\frac{\alpha}{\alpha+\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right)^r\left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\alpha+\exp(\boldsymbol{\beta}'\mathbf{x}_i)}\right)^y$$

- Known as the "Negbin II" model in most textbooks
- Collapses to the NBD when $\boldsymbol{\beta} = \mathbf{0}$

Fitting the NBD Regression Model

	A	B	C	D	E	F	G	H	I	J
1	r	0.1388			LL	-2888.966				
2	alpha	8.1979								
3	B_inc	0.0734								
4	B_sex	-0.0093								
5	B_age	0.9022								
6	B_size	-0.0243								
7				0.0734	-0.0093	0.9022	-0.0243			
8										
9	ID	Total		Income	Sex	Age	HH Size	exp(BX)	P(Y=y)	ln[P(Y=y)]
10	1	0		11.38	1	3.87	2	71.51161	0.72936	-0.316
11	2	5		9.77	1	4.04	1	76.02589	0.01587	-4.143
12	3	0						43.42559	0.77467	-0.255
13	4	0						72.50603	0.72810	-0.317
14	5	0		10.92	1					
15	6	0		10.92	0					
16	7	0		11.19	0					
17	8	1		11.74	0					
18	9	0		10.02	0	4.25	1	94.07931	0.70456	-0.350
19	10	0		10.92	0	3.85	3	66.80224	0.73555	-0.307
2736	2727	0		10.53	1	2.89	4	26.42093	0.81883	-0.200
2737	2728	0		11.74	1	2.83	3	28.08647	0.81351	-0.206

=EXP(SUMPRODUCT(D$7:G$7,D10:G10))

=EXP(GAMMALN(B$1+B10)-GAMMALN(B$1))/FACT(B10)*(B$2/(B$2+H10))^B$1*(H10/(B$2+H10))^B10

NBD Regression Results

Variable	Coefficient
r	0.1388
α	8.1979
Income	0.0734
Sex	−0.0093
Age	0.9022
HH Size	−0.0243
LL	−2889.0

S-Plus NBD Regression Results

```
Coefficients:
                  Value Std. Error    t value
(Intercept) -4.047149702 1.10159557 -3.6738979
     Income  0.074549233 0.09555222  0.7801936
        Sex -0.005240835 0.11592793 -0.0452077
        Age  0.889862966 0.14072030  6.3236289
    HH Size -0.025094493 0.04187696 -0.5992435

              Theta:  0.13878
          Std. Err.:  0.00726
```

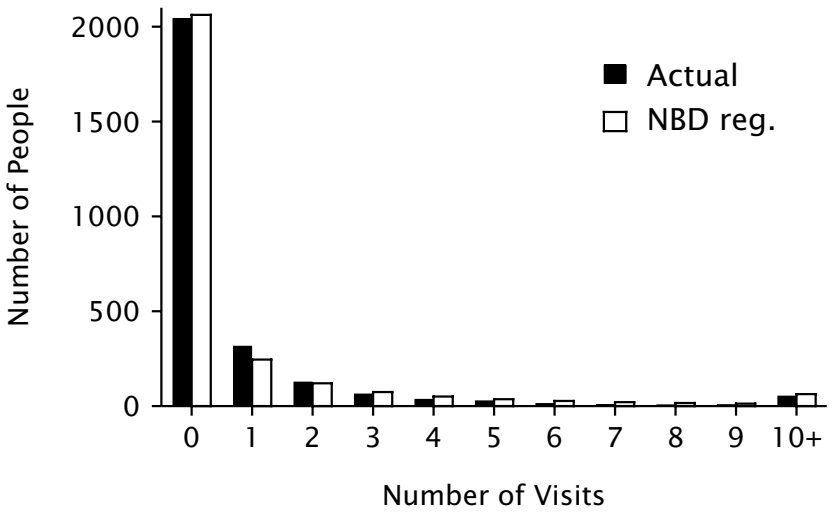

Limdep NBD Regression Results

Variable	Coefficient	Standard Error	b/St.Er.
Constant	-4.077239653	1.0451741	-3.901
INCOME	.7237686001E-01	.76663437E-01	.944
SEX	-.9009160129E-02	.11425700	-.079
AGE	.9045111135	.17741724	5.098
HH SIZE	-.2406546843E-01	.38695426E-01	-.622
	Overdispersion parameter		
Alpha	7.206708844	.33334006	21.620

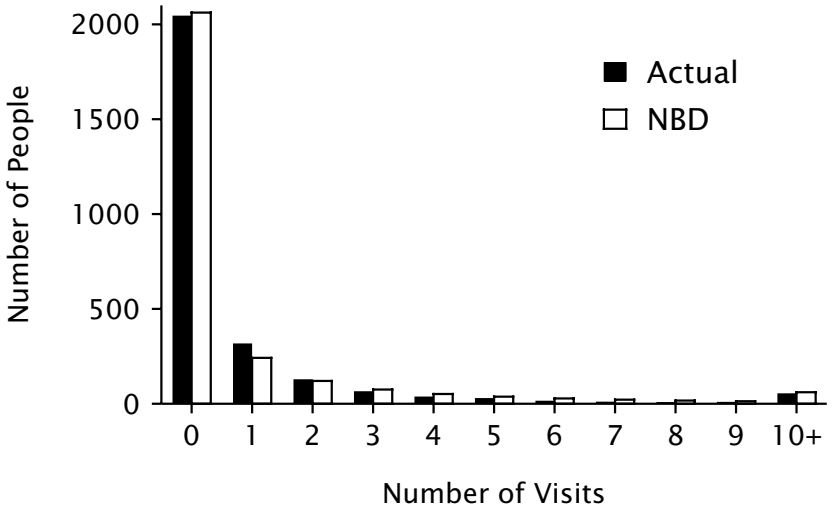
Summary of Regression Results

Variable	Poisson	ZIP	NBD
λ_0	0.0439	6.6231	
r			0.1388
α			8.1979
Income	0.0938	−0.0891	0.0734
Sex	0.0043	−0.1327	−0.0093
Age	0.5882	0.1141	0.9022
HH Size	−0.0359	0.0196	−0.0243
π		0.7433	
LL	−6291.5	−4297.5	−2889.0

Fit of the NBD Regression



Fit of the NBD



$\hat{r} = 0.134, \hat{\alpha} = 0.141, LL = -2905.6$

Concepts and Tools Introduced

- Incorporating covariate effects in count models
- Poisson (and NBD) regression models
- The possible over-emphasis of the value of covariates

Further Reading

Cameron, A. Colin and Pravin K. Trivedi (1998), *Regression Analysis of Count Data*, Cambridge: Cambridge University Press.

Wedel, Michel and Wagner A. Kamakura (2000), *Market Segmentation: Conceptual and Methodological Foundations*, 2nd edn., Boston, MA: Kluwer Academic Publishers.

Winkelmann, Rainer (2003), *Econometric Analysis of Count Data*, 4th edn., Berlin: Springer.

Introducing Covariates: The General Case

- Select a probability distribution that characterizes the individual-level behavior of interest:

$$f(y|\theta_i)$$

- Make the individual-level latent characteristic(s) a function of (time-invariant) covariates:

$$\theta_i = s(\theta_0, \mathbf{x}_i)$$

- Specify a mixing distribution to capture the heterogeneity in θ_i not "explained" by $\mathbf{x}_i$
- Derive the corresponding aggregate distribution

$$f(y|\mathbf{x}_i) = \int f(y|\theta_0, \mathbf{x}_i) g(\theta_0)\, d\theta_0$$

Covariates in Timing Models

- If the covariates are time-invariant, we can make λ a direct function of covariates:

$$F(t) = 1 - e^{-\lambda_0 \exp(\boldsymbol{\beta}' \mathbf{x}_i) t}$$

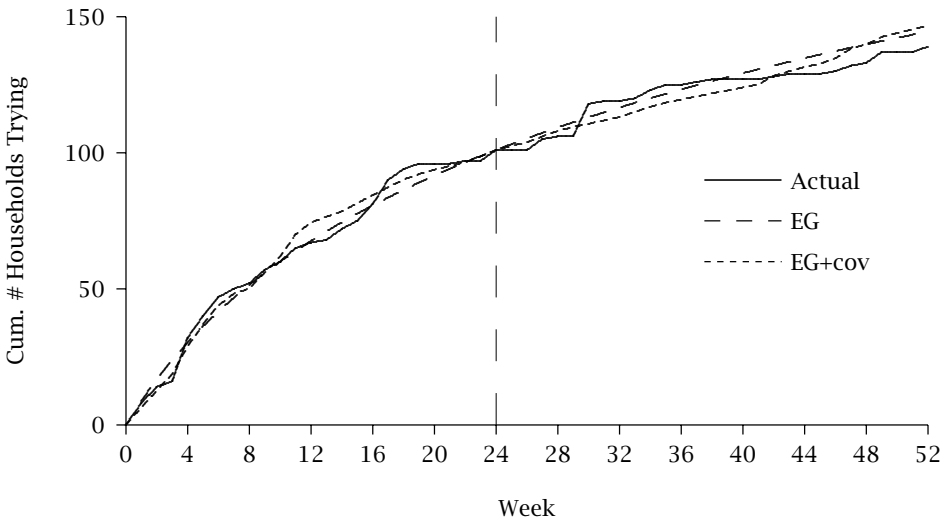
- If the covariates are time-varying (i.e., $\mathbf{x}_{it}$), we incorporate their effects via the hazard rate function

$$F(t) = 1 - e^{-\lambda_0 A(t)}$$

  where $A(t) = \sum_{j=1}^{t} \exp(\boldsymbol{\beta}' \mathbf{x}_{ij})$

- Known as "proportional hazards regression"

Comparing EG with EG+cov



Covariates in "Choice" Models

Two options for binary choice:

- The beta-logistic model
  - a generalization of the beta-binomial model in which the mean is made a function of (time-invariant) covariates
  - covariate effects not introduced at the level of the individual
- Finite mixture of binary logits:

$$P(Y = 1) = \frac{\exp(\boldsymbol{\beta}'\mathbf{x}_i)}{\exp(\boldsymbol{\beta}'\mathbf{x}_i) + 1}$$

  with some elements of $\boldsymbol{\beta}$ varying across segments

Discussion

Recap

- The preceding six 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.

Further Applications: Timing Models

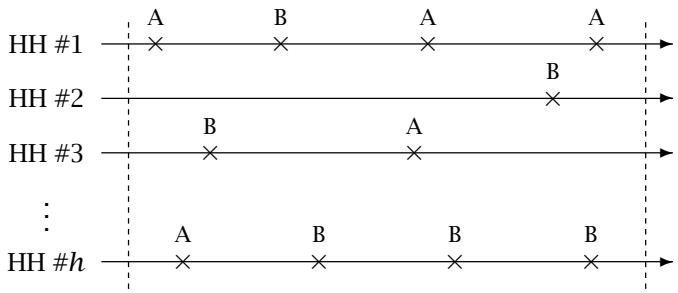
- Repeat purchasing of new products
- Response times:
  - Coupon redemptions
  - Survey response
  - Direct mail (response, returns, repeat sales)
- Other durations:
  - Salesforce job tenure
  - Length of web site browsing session

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
Stage 1	Counting			
	Timing			
	Choice			

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

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