

Applied Probability Models in Marketing Research: Introduction

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

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

13th Annual Advanced Research Techniques Forum
June 2-5, 2002

©2002 Bruce G. S. Hardie and Peter S. Fader

1

Problem 1: Predicting New Product Trial

(Modeling Timing Data)

2

Background

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

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

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

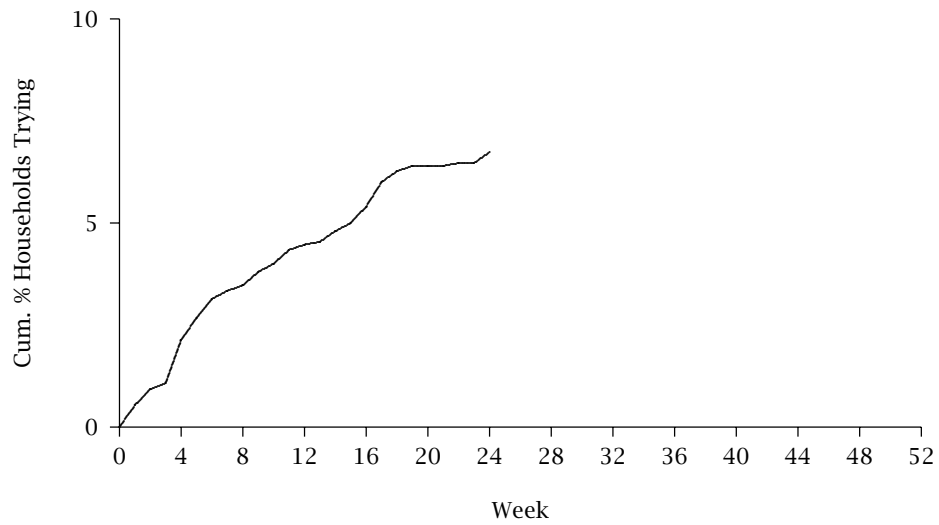
3

Krunchy Bits Cumulative Trial

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

4

Krunchy Bits Cumulative Trial



5

Approaches to Forecasting Trial

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

6

Developing a Model of Trial Purchasing

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

7

The Individual-Level Model

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

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

- λ represents the individual's trial rate.

8

The Market-Level Model

Assume two segments of consumers:

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

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

→ the “exponential w/ never triers” model

9

Estimating Model Parameters

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 $L(\cdot)$

10

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

11

Forecasting Trial

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

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

where N is the panel size.

- Use projection factors for market-level estimates.

12

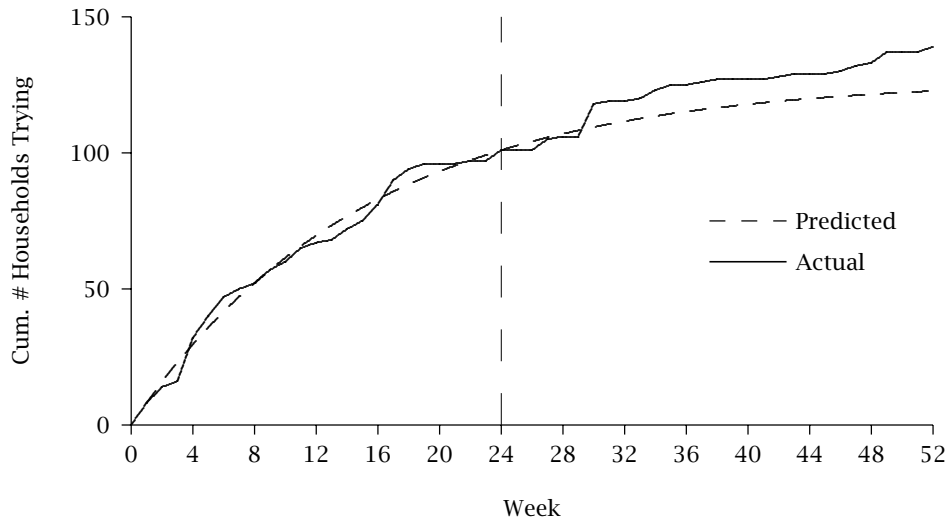
Problem 1 -- Model 1

	A	B	C	D	E	F	G	H
1	Product:	Krunchy Bits			p	0.5		
2	Panelists:	1499			\theta	0.05		
3					LL =	=SUM(F6:F30)		
4		Cum_Trl						
5	Week	# HHS	Incr_Trl		P(T <= t)			E[T(t)]
6	1	8	=B6		=F\$1*(1-EXP(-F\$2*A6))	=C6*LN(E6)		=B\$2*E6
7	2	14	=B7-B6		=F\$1*(1-EXP(-F\$2*A7))	=C7*LN(E7-E6)		=B\$2*E7
8	3	16	=B8-B7		=F\$1*(1-EXP(-F\$2*A8))	=C8*LN(E8-E7)		=B\$2*E8
9	4	32	=B9-B8		=F\$1*(1-EXP(-F\$2*A9))	=C9*LN(E9-E8)		=B\$2*E9
10	5	40	=B10-B9		=F\$1*(1-EXP(-F\$2*A10))	=C10*LN(E10-E9)		=B\$2*E10
11	6	47	=B11-B10		=F\$1*(1-EXP(-F\$2*A11))	=C11*LN(E11-E10)		=B\$2*E11
12	7	50	=B12-B11		=F\$1*(1-EXP(-F\$2*A12))	=C12*LN(E12-E11)		=B\$2*E12
13	8	52	=B13-B12		=F\$1*(1-EXP(-F\$2*A13))	=C13*LN(E13-E12)		=B\$2*E13
14	9	57	=B14-B13		=F\$1*(1-EXP(-F\$2*A14))	=C14*LN(E14-E13)		=B\$2*E14
15	10	60	=B15-B14		=F\$1*(1-EXP(-F\$2*A15))	=C15*LN(E15-E14)		=B\$2*E15
16	11	65	=B16-B15		=F\$1*(1-EXP(-F\$2*A16))	=C16*LN(E16-E15)		=B\$2*E16
17	12	67	=B17-B16		=F\$1*(1-EXP(-F\$2*A17))	=C17*LN(E17-E16)		=B\$2*E17
18	13	68	=B18-B17		=F\$1*(1-EXP(-F\$2*A18))	=C18*LN(E18-E17)		=B\$2*E18
19	14	72	=B19-B18		=F\$1*(1-EXP(-F\$2*A19))	=C19*LN(E19-E18)		=B\$2*E19
20	15	75	=B20-B19		=F\$1*(1-EXP(-F\$2*A20))	=C20*LN(E20-E19)		=B\$2*E20
21	16	81	=B21-B20		=F\$1*(1-EXP(-F\$2*A21))	=C21*LN(E21-E20)		=B\$2*E21
22	17	90	=B22-B21		=F\$1*(1-EXP(-F\$2*A22))	=C22*LN(E22-E21)		=B\$2*E22
23	18	94	=B23-B22		=F\$1*(1-EXP(-F\$2*A23))	=C23*LN(E23-E22)		=B\$2*E23
24	19	96	=B24-B23		=F\$1*(1-EXP(-F\$2*A24))	=C24*LN(E24-E23)		=B\$2*E24
25	20	96	=B25-B24		=F\$1*(1-EXP(-F\$2*A25))	=C25*LN(E25-E24)		=B\$2*E25
26	21	96	=B26-B25		=F\$1*(1-EXP(-F\$2*A26))	=C26*LN(E26-E25)		=B\$2*E26
27	22	97	=B27-B26		=F\$1*(1-EXP(-F\$2*A27))	=C27*LN(E27-E26)		=B\$2*E27
28	23	97	=B28-B27		=F\$1*(1-EXP(-F\$2*A28))	=C28*LN(E28-E27)		=B\$2*E28
29	24	101	=B29-B28		=F\$1*(1-EXP(-F\$2*A29))	=C29*LN(E29-E28)		=B\$2*E29
30	25	101			=F\$1*(1-EXP(-F\$2*A30))	=(B2-B29)*LN(1-E29)		=B\$2*E30
31	26	101			=F\$1*(1-EXP(-F\$2*A31))			=B\$2*E31
32	27	105			=F\$1*(1-EXP(-F\$2*A32))			=B\$2*E32
33	28	106			=F\$1*(1-EXP(-F\$2*A33))			=B\$2*E33
34	29	106			=F\$1*(1-EXP(-F\$2*A34))			=B\$2*E34
35	30	118			=F\$1*(1-EXP(-F\$2*A35))			=B\$2*E35
36	31	119			=F\$1*(1-EXP(-F\$2*A36))			=B\$2*E36
37	32	119			=F\$1*(1-EXP(-F\$2*A37))			=B\$2*E37
38	33	120			=F\$1*(1-EXP(-F\$2*A38))			=B\$2*E38
39	34	123			=F\$1*(1-EXP(-F\$2*A39))			=B\$2*E39
40	35	125			=F\$1*(1-EXP(-F\$2*A40))			=B\$2*E40
41	36	125			=F\$1*(1-EXP(-F\$2*A41))			=B\$2*E41
42	37	126			=F\$1*(1-EXP(-F\$2*A42))			=B\$2*E42
43	38	127			=F\$1*(1-EXP(-F\$2*A43))			=B\$2*E43
44	39	127			=F\$1*(1-EXP(-F\$2*A44))			=B\$2*E44
45	40	127			=F\$1*(1-EXP(-F\$2*A45))			=B\$2*E45
46	41	127			=F\$1*(1-EXP(-F\$2*A46))			=B\$2*E46
47	42	128			=F\$1*(1-EXP(-F\$2*A47))			=B\$2*E47
48	43	129			=F\$1*(1-EXP(-F\$2*A48))			=B\$2*E48
49	44	129			=F\$1*(1-EXP(-F\$2*A49))			=B\$2*E49
50	45	129			=F\$1*(1-EXP(-F\$2*A50))			=B\$2*E50
51	46	130			=F\$1*(1-EXP(-F\$2*A51))			=B\$2*E51
52	47	132			=F\$1*(1-EXP(-F\$2*A52))			=B\$2*E52
53	48	133			=F\$1*(1-EXP(-F\$2*A53))			=B\$2*E53
54	49	137			=F\$1*(1-EXP(-F\$2*A54))			=B\$2*E54
55	50	137			=F\$1*(1-EXP(-F\$2*A55))			=B\$2*E55
56	51	137			=F\$1*(1-EXP(-F\$2*A56))			=B\$2*E56
57	52	139			=F\$1*(1-EXP(-F\$2*A57))			=B\$2*E57

Problem 1 -- Model 1

	A	B	C	D	E	F	G	H
1	Product:	Krunchy Bits			p	0.08456		
2	Panelists:	1499			\theta	0.0664		
3					LL =	-680.9094		
4		Cum_Trl						
5	Week	# HHs	Incr_Trl		P(T <= t)			E[T(t)]
6	1	8	8		0.00543	-41.723		8.14
7	2	14	6		0.01052	-31.691		15.76
8	3	16	2		0.01527	-10.696		22.89
9	4	32	16		0.01972	-86.633		29.57
10	5	40	8		0.02389	-43.848		35.81
11	6	47	7		0.02779	-38.832		41.65
12	7	50	3		0.03143	-16.841		47.12
13	8	52	2		0.03485	-11.360		52.24
14	9	57	5		0.03804	-28.733		57.02
15	10	60	3		0.04103	-17.439		61.50
16	11	65	5		0.04383	-29.397		65.70
17	12	67	2		0.04644	-11.892		69.62
18	13	68	1		0.04889	-6.012		73.29
19	14	72	4		0.05118	-24.314		76.72
20	15	75	3		0.05333	-18.435		79.94
21	16	81	6		0.05533	-37.268		82.95
22	17	90	9		0.05721	-56.500		85.76
23	18	94	4		0.05897	-25.377		88.39
24	19	96	2		0.06061	-12.821		90.86
25	20	96	0		0.06215	0.000		93.16
26	21	96	0		0.06359	0.000		95.32
27	22	97	1		0.06494	-6.610		97.34
28	23	97	0		0.06620	0.000		99.23
29	24	101	4		0.06738	-26.970		101.00
30	25	101			0.06848	-97.518		102.65
31	26	101			0.06951			104.20
32	27	105			0.07048			105.65
33	28	106			0.07139			107.01
34	29	106			0.07223			108.28
35	30	118			0.07302			109.46
36	31	119			0.07377			110.57
37	32	119			0.07446			111.61
38	33	120			0.07511			112.59
39	34	123			0.07572			113.50
40	35	125			0.07628			114.35
41	36	125			0.07682			115.15
42	37	126			0.07731			115.89
43	38	127			0.07778			116.59
44	39	127			0.07821			117.24
45	40	127			0.07862			117.85
46	41	127			0.07900			118.43
47	42	128			0.07936			118.96
48	43	129			0.07969			119.46
49	44	129			0.08001			119.93
50	45	129			0.08030			120.37
51	46	130			0.08057			120.78
52	47	132			0.08083			121.16
53	48	133			0.08107			121.52
54	49	137			0.08129			121.86
55	50	137			0.08150			122.17
56	51	137			0.08170			122.47
57	52	139			0.08188			122.74

Cumulative Trial Forecast



13

Extending the Basic Model

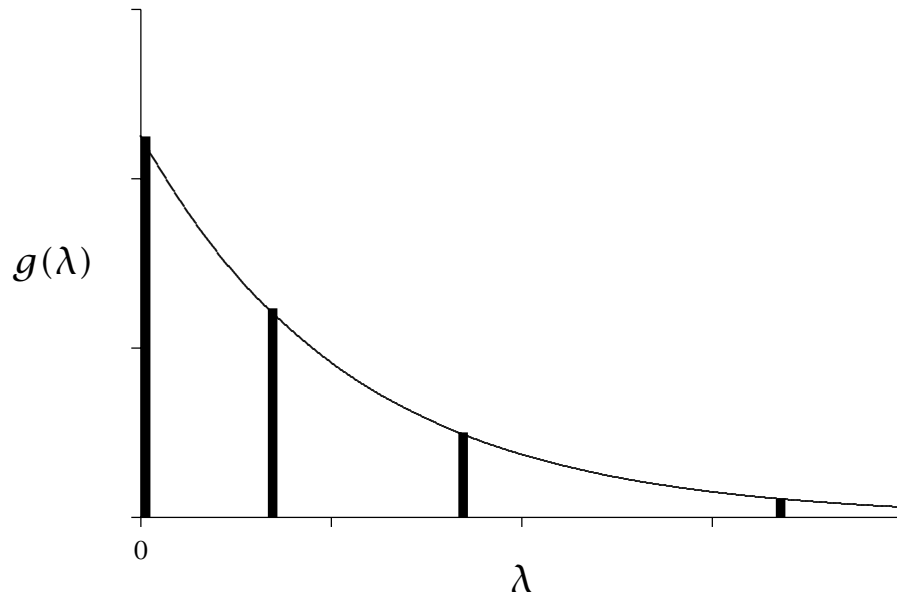
- The “exponential w/ never triers” model assumes all triers have the same underlying trial rate θ — a bit simplistic.
- Allow for multiple trier “segments” each with a different (latent) trial rate:

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

- Replace the discrete distribution with a continuous distribution.

14

Distribution of Trial Rates



15

Distribution of Trial Rates

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

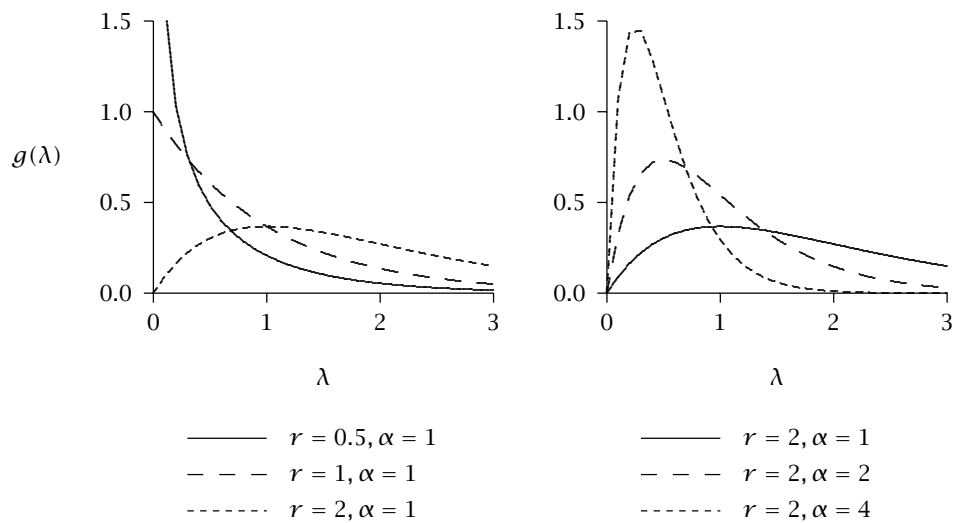
$$g(\lambda) = \frac{\alpha^r \lambda^{r-1} e^{-\alpha\lambda}}{\Gamma(r)}$$

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

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

16

Illustrative Gamma Density Functions



17

Alternative Market-Level Model

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

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

We call this the “exponential-gamma” model.

18

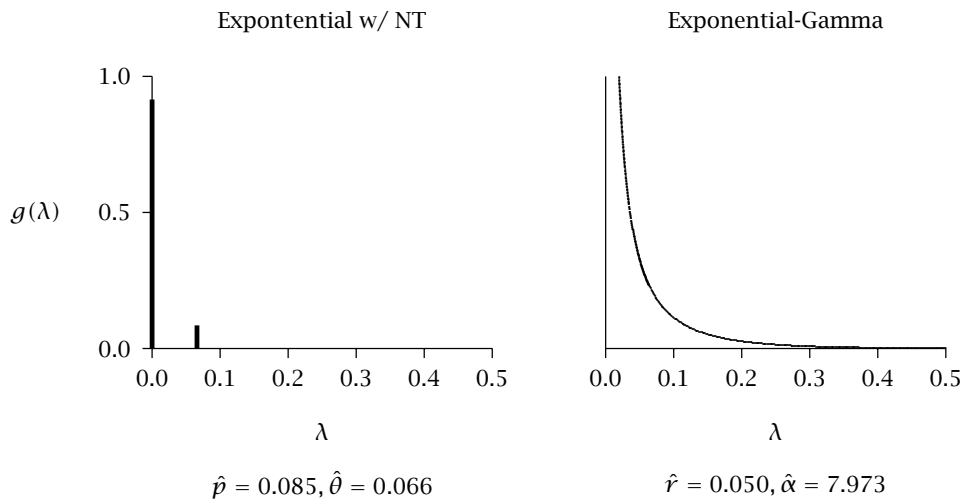
Estimating Model Parameters

The log-likelihood function is defined as:

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

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

Estimated Distribution of λ



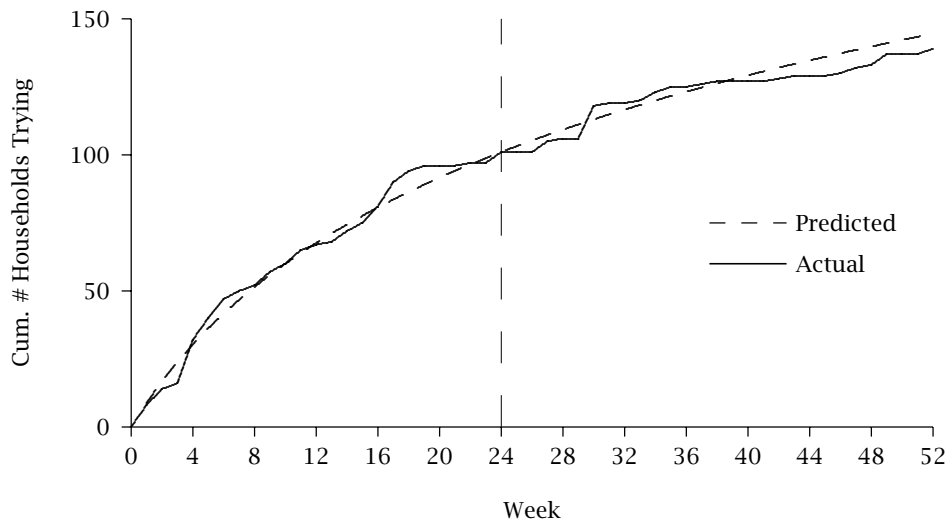
Problem 1 -- Model 2

	A	B	C	D	E	F	G	H
1	Product:	Krunchy Bits			r	1		
2	Panelists:	1499			\alpha	1		
3					LL =	=SUM(F6:F30)		
4		Cum_Trl						
5	Week	# HHS	Incr_Trl		P(T <= t)			E[T(t)]
6	1	8	=B6		=1-(F\$2/(F\$2+A6))^F\$1	=C6*LN(E6)		=B\$2*E6
7	2	14	=B7-B6		=1-(F\$2/(F\$2+A7))^F\$1	=C7*LN(E7-E6)		=B\$2*E7
8	3	16	=B8-B7		=1-(F\$2/(F\$2+A8))^F\$1	=C8*LN(E8-E7)		=B\$2*E8
9	4	32	=B9-B8		=1-(F\$2/(F\$2+A9))^F\$1	=C9*LN(E9-E8)		=B\$2*E9
10	5	40	=B10-B9		=1-(F\$2/(F\$2+A10))^F\$1	=C10*LN(E10-E9)		=B\$2*E10
11	6	47	=B11-B10		=1-(F\$2/(F\$2+A11))^F\$1	=C11*LN(E11-E10)		=B\$2*E11
12	7	50	=B12-B11		=1-(F\$2/(F\$2+A12))^F\$1	=C12*LN(E12-E11)		=B\$2*E12
13	8	52	=B13-B12		=1-(F\$2/(F\$2+A13))^F\$1	=C13*LN(E13-E12)		=B\$2*E13
14	9	57	=B14-B13		=1-(F\$2/(F\$2+A14))^F\$1	=C14*LN(E14-E13)		=B\$2*E14
15	10	60	=B15-B14		=1-(F\$2/(F\$2+A15))^F\$1	=C15*LN(E15-E14)		=B\$2*E15
16	11	65	=B16-B15		=1-(F\$2/(F\$2+A16))^F\$1	=C16*LN(E16-E15)		=B\$2*E16
17	12	67	=B17-B16		=1-(F\$2/(F\$2+A17))^F\$1	=C17*LN(E17-E16)		=B\$2*E17
18	13	68	=B18-B17		=1-(F\$2/(F\$2+A18))^F\$1	=C18*LN(E18-E17)		=B\$2*E18
19	14	72	=B19-B18		=1-(F\$2/(F\$2+A19))^F\$1	=C19*LN(E19-E18)		=B\$2*E19
20	15	75	=B20-B19		=1-(F\$2/(F\$2+A20))^F\$1	=C20*LN(E20-E19)		=B\$2*E20
21	16	81	=B21-B20		=1-(F\$2/(F\$2+A21))^F\$1	=C21*LN(E21-E20)		=B\$2*E21
22	17	90	=B22-B21		=1-(F\$2/(F\$2+A22))^F\$1	=C22*LN(E22-E21)		=B\$2*E22
23	18	94	=B23-B22		=1-(F\$2/(F\$2+A23))^F\$1	=C23*LN(E23-E22)		=B\$2*E23
24	19	96	=B24-B23		=1-(F\$2/(F\$2+A24))^F\$1	=C24*LN(E24-E23)		=B\$2*E24
25	20	96	=B25-B24		=1-(F\$2/(F\$2+A25))^F\$1	=C25*LN(E25-E24)		=B\$2*E25
26	21	96	=B26-B25		=1-(F\$2/(F\$2+A26))^F\$1	=C26*LN(E26-E25)		=B\$2*E26
27	22	97	=B27-B26		=1-(F\$2/(F\$2+A27))^F\$1	=C27*LN(E27-E26)		=B\$2*E27
28	23	97	=B28-B27		=1-(F\$2/(F\$2+A28))^F\$1	=C28*LN(E28-E27)		=B\$2*E28
29	24	101	=B29-B28		=1-(F\$2/(F\$2+A29))^F\$1	=C29*LN(E29-E28)		=B\$2*E29
30	25	101			=1-(F\$2/(F\$2+A30))^F\$1	=(B2-B29)*LN(1-E29)		=B\$2*E30
31	26	101			=1-(F\$2/(F\$2+A31))^F\$1			=B\$2*E31
32	27	105			=1-(F\$2/(F\$2+A32))^F\$1			=B\$2*E32
33	28	106			=1-(F\$2/(F\$2+A33))^F\$1			=B\$2*E33
34	29	106			=1-(F\$2/(F\$2+A34))^F\$1			=B\$2*E34
35	30	118			=1-(F\$2/(F\$2+A35))^F\$1			=B\$2*E35
36	31	119			=1-(F\$2/(F\$2+A36))^F\$1			=B\$2*E36
37	32	119			=1-(F\$2/(F\$2+A37))^F\$1			=B\$2*E37
38	33	120			=1-(F\$2/(F\$2+A38))^F\$1			=B\$2*E38
39	34	123			=1-(F\$2/(F\$2+A39))^F\$1			=B\$2*E39
40	35	125			=1-(F\$2/(F\$2+A40))^F\$1			=B\$2*E40
41	36	125			=1-(F\$2/(F\$2+A41))^F\$1			=B\$2*E41
42	37	126			=1-(F\$2/(F\$2+A42))^F\$1			=B\$2*E42
43	38	127			=1-(F\$2/(F\$2+A43))^F\$1			=B\$2*E43
44	39	127			=1-(F\$2/(F\$2+A44))^F\$1			=B\$2*E44
45	40	127			=1-(F\$2/(F\$2+A45))^F\$1			=B\$2*E45
46	41	127			=1-(F\$2/(F\$2+A46))^F\$1			=B\$2*E46
47	42	128			=1-(F\$2/(F\$2+A47))^F\$1			=B\$2*E47
48	43	129			=1-(F\$2/(F\$2+A48))^F\$1			=B\$2*E48
49	44	129			=1-(F\$2/(F\$2+A49))^F\$1			=B\$2*E49
50	45	129			=1-(F\$2/(F\$2+A50))^F\$1			=B\$2*E50
51	46	130			=1-(F\$2/(F\$2+A51))^F\$1			=B\$2*E51
52	47	132			=1-(F\$2/(F\$2+A52))^F\$1			=B\$2*E52
53	48	133			=1-(F\$2/(F\$2+A53))^F\$1			=B\$2*E53
54	49	137			=1-(F\$2/(F\$2+A54))^F\$1			=B\$2*E54
55	50	137			=1-(F\$2/(F\$2+A55))^F\$1			=B\$2*E55
56	51	137			=1-(F\$2/(F\$2+A56))^F\$1			=B\$2*E56
57	52	139			=1-(F\$2/(F\$2+A57))^F\$1			=B\$2*E57

Problem 1 -- Model 2

	A	B	C	D	E	F	G	H
1	Product:	Krunchy Bits			r	0.050245		
2	Panelists:	1499			\alpha	7.973267		
3					LL =	-681.3729		
4		Cum_Trl						
5	Week	# HHs	Incr_Trl		P(T <= t)			E[T(t)]
6	1	8	8		0.00592	-41.036		8.87
7	2	14	6		0.01118	-31.482		16.76
8	3	16	2		0.01592	-10.705		23.86
9	4	32	16		0.02022	-87.175		30.31
10	5	40	8		0.02416	-44.291		36.22
11	6	47	7		0.02780	-39.322		41.67
12	7	50	3		0.03117	-17.078		46.72
13	8	52	2		0.03431	-11.526		51.43
14	9	57	5		0.03725	-29.144		55.84
15	10	60	3		0.04002	-17.672		59.98
16	11	65	5		0.04262	-29.746		63.89
17	12	67	2		0.04509	-12.009		67.59
18	13	68	1		0.04743	-6.057		71.10
19	14	72	4		0.04966	-24.429		74.44
20	15	75	3		0.05178	-18.465		77.62
21	16	81	6		0.05381	-37.205		80.66
22	17	90	9		0.05575	-56.202		83.57
23	18	94	4		0.05761	-25.147		86.36
24	19	96	2		0.05940	-12.654		89.04
25	20	96	0		0.06112	0.000		91.62
26	21	96	0		0.06277	0.000		94.10
27	22	97	1		0.06437	-6.440		96.49
28	23	97	0		0.06591	0.000		98.80
29	24	101	4		0.06740	-26.036		101.04
30	25	101			0.06884	-97.554		103.20
31	26	101			0.07024			105.29
32	27	105			0.07159			107.32
33	28	106			0.07291			109.29
34	29	106			0.07419			111.20
35	30	118			0.07543			113.06
36	31	119			0.07663			114.87
37	32	119			0.07781			116.63
38	33	120			0.07895			118.35
39	34	123			0.08007			120.02
40	35	125			0.08115			121.65
41	36	125			0.08222			123.24
42	37	126			0.08325			124.80
43	38	127			0.08426			126.31
44	39	127			0.08525			127.80
45	40	127			0.08622			129.25
46	41	127			0.08717			130.67
47	42	128			0.08810			132.05
48	43	129			0.08900			133.42
49	44	129			0.08989			134.75
50	45	129			0.09076			136.05
51	46	130			0.09162			137.33
52	47	132			0.09245			138.59
53	48	133			0.09328			139.82
54	49	137			0.09408			141.03
55	50	137			0.09487			142.22
56	51	137			0.09565			143.38
57	52	139			0.09641			144.53

Cumulative Trial Forecast



21

Further Model Extensions

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

22

Concepts and Tools Introduced

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

23

Further Reading

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

Fader, Peter S., Bruce G. S. Hardie, and Robert Zeithammer (2002), "Forecasting New Product Trial in a Controlled Test Market Environment," unpublished working paper. (Available at <http://brucehardie.com/>)

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

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

24

Introduction to Probability Models

25

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.

26

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

27

Building a Probability Model

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

28

Building a Probability Model

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

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

29

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.

30

Outline

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

31

Problem 2: Estimating Billboard Exposures (Modeling Count Data)

32

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.

33

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.

34

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.

35

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

36

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.

37

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).

38

Modeling One Week Exposures

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

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

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

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

39

Modeling One Week Exposures

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

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

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

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

40

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:

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

41

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

42

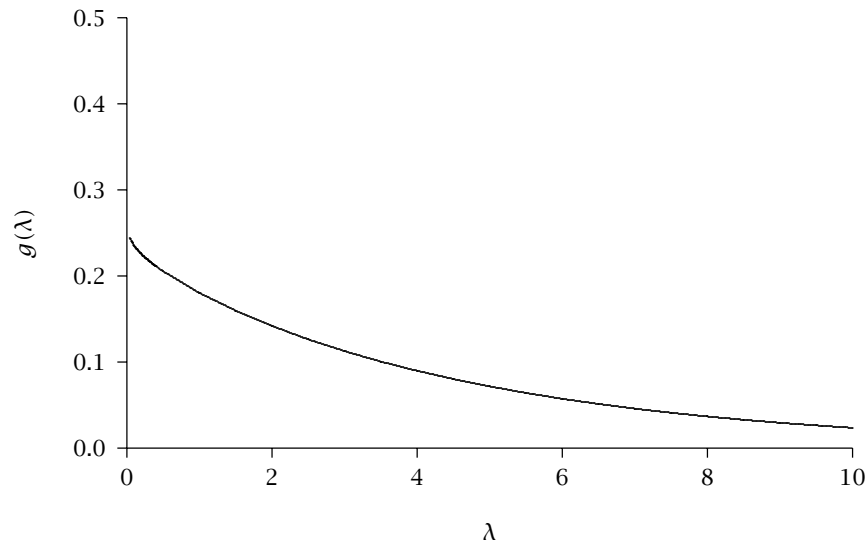
Problem 2 -- Parameter Estimation

	A	B	C	D
1	r	1		
2	\alpha	1		LL= =SUM(D5:D28)
3				
4	x	f_x		P(X=x)
5	0	48	=B2/(B2+1)^B1	=B5*LN(C5)
6	1	37	=C5*(B\$1+A6-1)/(A6*(B\$2+1))	=B6*LN(C6)
7	2	30	=C6*(B\$1+A7-1)/(A7*(B\$2+1))	=B7*LN(C7)
8	3	24	=C7*(B\$1+A8-1)/(A8*(B\$2+1))	=B8*LN(C8)
9	4	20	=C8*(B\$1+A9-1)/(A9*(B\$2+1))	=B9*LN(C9)
10	5	16	=C9*(B\$1+A10-1)/(A10*(B\$2+1))	=B10*LN(C10)
11	6	13	=C10*(B\$1+A11-1)/(A11*(B\$2+1))	=B11*LN(C11)
12	7	11	=C11*(B\$1+A12-1)/(A12*(B\$2+1))	=B12*LN(C12)
13	8	9	=C12*(B\$1+A13-1)/(A13*(B\$2+1))	=B13*LN(C13)
14	9	7	=C13*(B\$1+A14-1)/(A14*(B\$2+1))	=B14*LN(C14)
15	10	6	=C14*(B\$1+A15-1)/(A15*(B\$2+1))	=B15*LN(C15)
16	11	5	=C15*(B\$1+A16-1)/(A16*(B\$2+1))	=B16*LN(C16)
17	12	5	=C16*(B\$1+A17-1)/(A17*(B\$2+1))	=B17*LN(C17)
18	13	3	=C17*(B\$1+A18-1)/(A18*(B\$2+1))	=B18*LN(C18)
19	14	3	=C18*(B\$1+A19-1)/(A19*(B\$2+1))	=B19*LN(C19)
20	15	2	=C19*(B\$1+A20-1)/(A20*(B\$2+1))	=B20*LN(C20)
21	16	2	=C20*(B\$1+A21-1)/(A21*(B\$2+1))	=B21*LN(C21)
22	17	2	=C21*(B\$1+A22-1)/(A22*(B\$2+1))	=B22*LN(C22)
23	18	1	=C22*(B\$1+A23-1)/(A23*(B\$2+1))	=B23*LN(C23)
24	19	1	=C23*(B\$1+A24-1)/(A24*(B\$2+1))	=B24*LN(C24)
25	20	2	=C24*(B\$1+A25-1)/(A25*(B\$2+1))	=B25*LN(C25)
26	21	1	=C25*(B\$1+A26-1)/(A26*(B\$2+1))	=B26*LN(C26)
27	22	1	=C26*(B\$1+A27-1)/(A27*(B\$2+1))	=B27*LN(C27)
28	23	1	=C27*(B\$1+A28-1)/(A28*(B\$2+1))	=B28*LN(C28)

Problem 2 -- Parameter Estimation

	A	B	C	D
1	r	0.96926		
2	\alpha	0.21752	LL=	-649.6888
3				
4	x	f_x	P(X=x)	
5	0	48	0.18837	-80.128
6	1	37	0.14996	-70.203
7	2	30	0.12128	-63.291
8	3	24	0.09859	-55.603
9	4	20	0.08035	-50.427
10	5	16	0.06559	-43.589
11	6	13	0.05360	-38.041
12	7	11	0.04383	-34.402
13	8	9	0.03586	-29.953
14	9	7	0.02935	-24.699
15	10	6	0.02403	-22.370
16	11	5	0.01969	-19.639
17	12	5	0.01613	-20.636
18	13	3	0.01321	-12.979
19	14	3	0.01083	-13.576
20	15	2	0.00888	-9.449
21	16	2	0.00728	-9.846
22	17	2	0.00597	-10.243
23	18	1	0.00489	-5.320
24	19	1	0.00401	-5.519
25	20	2	0.00329	-11.434
26	21	1	0.00270	-5.915
27	22	1	0.00221	-6.113
28	23	1	0.00182	-6.312

Estimated Distribution of λ



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

43

NBD for a Non-Unit Time Period

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

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

44

NBD for a Non-Unit Time Period

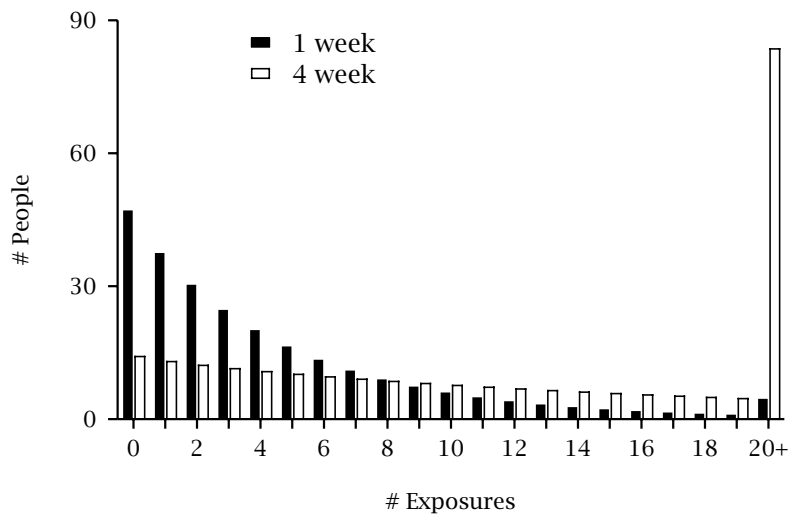
- The distribution of exposures at the population-level is given by:

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

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

45

Exposure Distributions: 1 week vs. 4 week



46

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

47

Concepts and Tools Introduced

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

48

Problem 2 -- Solution

	A	B
1	r	=Parameter Estimation!B1
2	\alpha	=Parameter Estimation!B2
3	t	4
4		
5	$P(X(t)=0)$	$=(B2/(B2+B3))^{B1}$
6	$E[X(t)]$	$=B1*B3/B2$
7		
8	Reach	$=1-B5$
9	Frequency	$=B6/B8$
10	GRPs	$=100*B6$

Problem 2 -- Solution

	A	B
1	r	0.96926
2	\alpha	0.21752
3	t	4
4		
5	$P(X(t)=0)$	0.056
6	$E[X(t)]$	17.82
7		
8	Reach	94.4%
9	Frequency	18.9
10	GRPs	1782

Further Reading

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

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

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

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

The “Segmentation” Approach

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

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

51

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

52

Ben's Knick Knacks, Inc.

Standard approach:

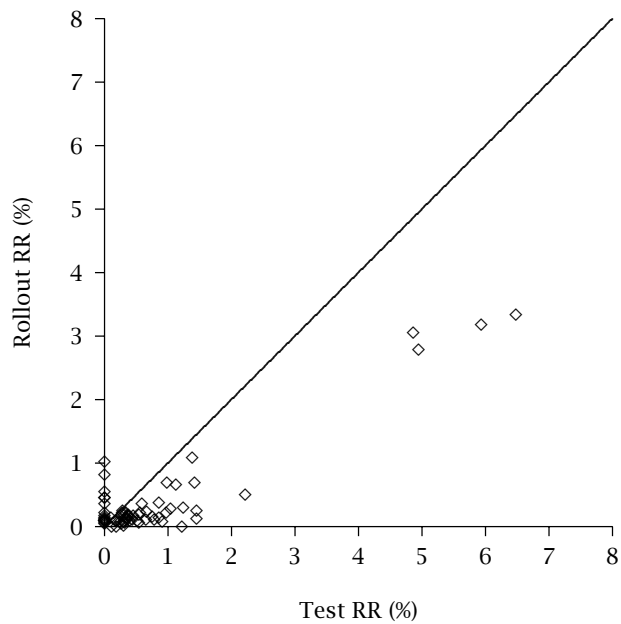
- Rollout to all segments with

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

- 51 segments pass this hurdle

53

Test vs. Actual Response Rate



54

Modeling Objective

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

55

Model Development

Notation:

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

m_s = # members of segment s tested

X_s = # responses to test in segment s

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

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

56

Distribution of Response Probabilities

- Heterogeneity in p_s is captured using a beta distribution:

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

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

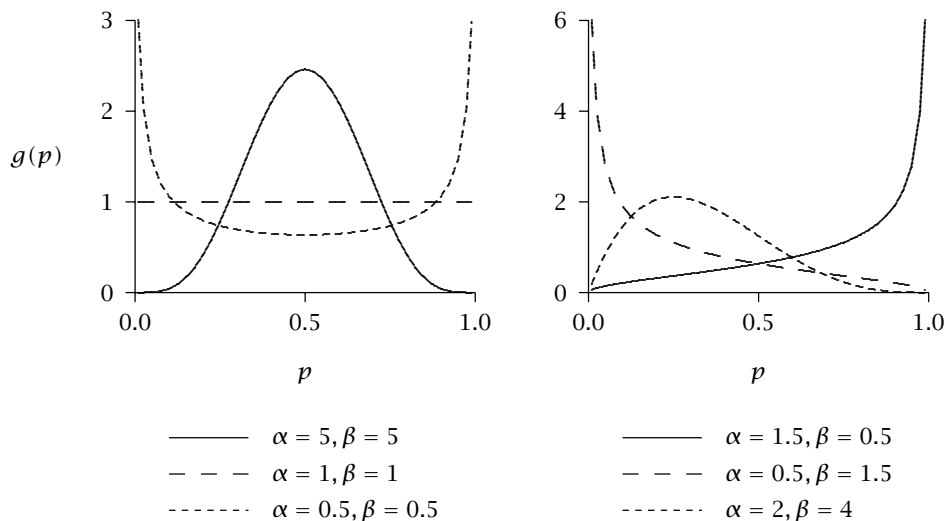
$$B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha + \beta)}$$

- The mean of the beta distribution is given by

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

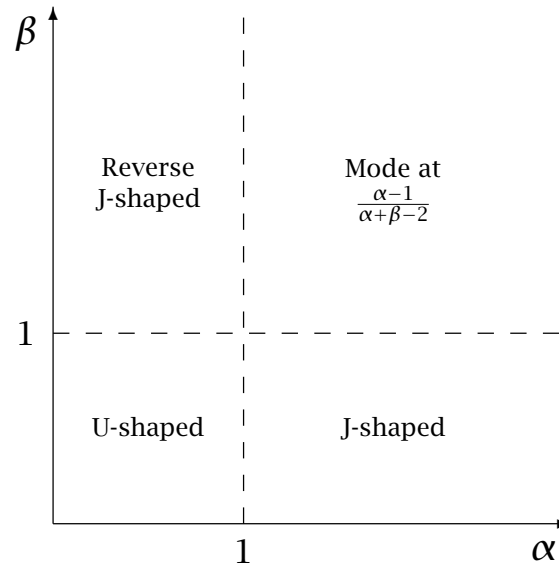
57

Illustrative Beta Density Functions



58

Shape of the Beta Density



59

The Beta Binomial Model

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

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

60

Estimating Model Parameters

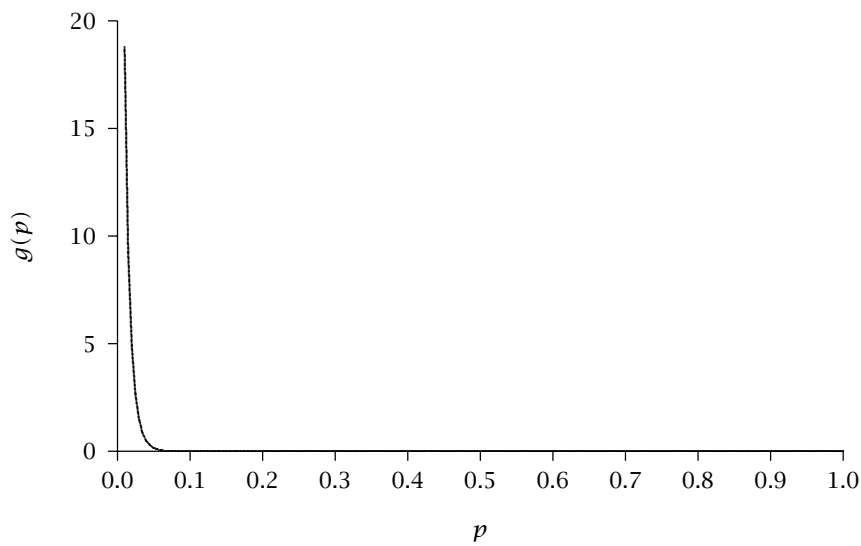
The log-likelihood function is defined as:

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

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

61

Estimated Distribution of p



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

62

Problem 3 -- Model (a)

	A	B	C	D	E	F
1	\alpha	1		B(\alpha, \beta)	=EXP(GAMMALN(B1)+GAMMALN(B2)-GAMMALN(B1+B2))	
2	\beta	1				
3					LL =	=SUM(F6:F131)
4						
5	Segment	m_s	x_s		P(X=x m)	
6	1	34	0		=COMBIN(B6,C6)*EXP(GAMMALN(B\$1+C6)+GAMMALN(B\$2+B6-C6)-GAMMALN(B\$1+B\$2+B6))/E\$1	=LN(E6)
7	2	102	1		=COMBIN(B7,C7)*EXP(GAMMALN(B\$1+C7)+GAMMALN(B\$2+B7-C7)-GAMMALN(B\$1+B\$2+B7))/E\$1	=LN(E7)
8	3	53	0		=COMBIN(B8,C8)*EXP(GAMMALN(B\$1+C8)+GAMMALN(B\$2+B8-C8)-GAMMALN(B\$1+B\$2+B8))/E\$1	=LN(E8)
9	4	145	2		=COMBIN(B9,C9)*EXP(GAMMALN(B\$1+C9)+GAMMALN(B\$2+B9-C9)-GAMMALN(B\$1+B\$2+B9))/E\$1	=LN(E9)
10	5	1254	62		=COMBIN(B10,C10)*EXP(GAMMALN(B\$1+C10)+GAMMALN(B\$2+B10-C10)-GAMMALN(B\$1+B\$2+B10))/E\$1	=LN(E10)
11	6	144	7		=COMBIN(B11,C11)*EXP(GAMMALN(B\$1+C11)+GAMMALN(B\$2+B11-C11)-GAMMALN(B\$1+B\$2+B11))/E\$1	=LN(E11)
12	7	1235	80		=COMBIN(B12,C12)*EXP(GAMMALN(B\$1+C12)+GAMMALN(B\$2+B12-C12)-GAMMALN(B\$1+B\$2+B12))/E\$1	=LN(E12)
13	8	573	34		=COMBIN(B13,C13)*EXP(GAMMALN(B\$1+C13)+GAMMALN(B\$2+B13-C13)-GAMMALN(B\$1+B\$2+B13))/E\$1	=LN(E13)
14	9	1083	24		=COMBIN(B14,C14)*EXP(GAMMALN(B\$1+C14)+GAMMALN(B\$2+B14-C14)-GAMMALN(B\$1+B\$2+B14))/E\$1	=LN(E14)
15	10	352	5		=COMBIN(B15,C15)*EXP(GAMMALN(B\$1+C15)+GAMMALN(B\$2+B15-C15)-GAMMALN(B\$1+B\$2+B15))/E\$1	=LN(E15)
16	11	817	7		=COMBIN(B16,C16)*EXP(GAMMALN(B\$1+C16)+GAMMALN(B\$2+B16-C16)-GAMMALN(B\$1+B\$2+B16))/E\$1	=LN(E16)
17	12	118	0		=COMBIN(B17,C17)*EXP(GAMMALN(B\$1+C17)+GAMMALN(B\$2+B17-C17)-GAMMALN(B\$1+B\$2+B17))/E\$1	=LN(E17)
18	13	1049	3		=COMBIN(B18,C18)*EXP(GAMMALN(B\$1+C18)+GAMMALN(B\$2+B18-C18)-GAMMALN(B\$1+B\$2+B18))/E\$1	=LN(E18)
19	14	452	3		=COMBIN(B19,C19)*EXP(GAMMALN(B\$1+C19)+GAMMALN(B\$2+B19-C19)-GAMMALN(B\$1+B\$2+B19))/E\$1	=LN(E19)
20	15	338	2		=COMBIN(B20,C20)*EXP(GAMMALN(B\$1+C20)+GAMMALN(B\$2+B20-C20)-GAMMALN(B\$1+B\$2+B20))/E\$1	=LN(E20)
21	16	168	0		=COMBIN(B21,C21)*EXP(GAMMALN(B\$1+C21)+GAMMALN(B\$2+B21-C21)-GAMMALN(B\$1+B\$2+B21))/E\$1	=LN(E21)
22	17	242	3		=COMBIN(B22,C22)*EXP(GAMMALN(B\$1+C22)+GAMMALN(B\$2+B22-C22)-GAMMALN(B\$1+B\$2+B22))/E\$1	=LN(E22)
23	18	185	1		=COMBIN(B23,C23)*EXP(GAMMALN(B\$1+C23)+GAMMALN(B\$2+B23-C23)-GAMMALN(B\$1+B\$2+B23))/E\$1	=LN(E23)
24	19	116	0		=COMBIN(B24,C24)*EXP(GAMMALN(B\$1+C24)+GAMMALN(B\$2+B24-C24)-GAMMALN(B\$1+B\$2+B24))/E\$1	=LN(E24)
25	20	69	1		=COMBIN(B25,C25)*EXP(GAMMALN(B\$1+C25)+GAMMALN(B\$2+B25-C25)-GAMMALN(B\$1+B\$2+B25))/E\$1	=LN(E25)
26	21	193	1		=COMBIN(B26,C26)*EXP(GAMMALN(B\$1+C26)+GAMMALN(B\$2+B26-C26)-GAMMALN(B\$1+B\$2+B26))/E\$1	=LN(E26)
27	22	82	1		=COMBIN(B27,C27)*EXP(GAMMALN(B\$1+C27)+GAMMALN(B\$2+B27-C27)-GAMMALN(B\$1+B\$2+B27))/E\$1	=LN(E27)
28	23	265	1		=COMBIN(B28,C28)*EXP(GAMMALN(B\$1+C28)+GAMMALN(B\$2+B28-C28)-GAMMALN(B\$1+B\$2+B28))/E\$1	=LN(E28)
29	24	171	0		=COMBIN(B29,C29)*EXP(GAMMALN(B\$1+C29)+GAMMALN(B\$2+B29-C29)-GAMMALN(B\$1+B\$2+B29))/E\$1	=LN(E29)
30	25	1554	7		=COMBIN(B30,C30)*EXP(GAMMALN(B\$1+C30)+GAMMALN(B\$2+B30-C30)-GAMMALN(B\$1+B\$2+B30))/E\$1	=LN(E30)
31	26	1339	4		=COMBIN(B31,C31)*EXP(GAMMALN(B\$1+C31)+GAMMALN(B\$2+B31-C31)-GAMMALN(B\$1+B\$2+B31))/E\$1	=LN(E31)
32	27	1167	4		=COMBIN(B32,C32)*EXP(GAMMALN(B\$1+C32)+GAMMALN(B\$2+B32-C32)-GAMMALN(B\$1+B\$2+B32))/E\$1	=LN(E32)
33	28	621	2		=COMBIN(B33,C33)*EXP(GAMMALN(B\$1+C33)+GAMMALN(B\$2+B33-C33)-GAMMALN(B\$1+B\$2+B33))/E\$1	=LN(E33)
34	29	1013	1		=COMBIN(B34,C34)*EXP(GAMMALN(B\$1+C34)+GAMMALN(B\$2+B34-C34)-GAMMALN(B\$1+B\$2+B34))/E\$1	=LN(E34)
35	30	544	1		=COMBIN(B35,C35)*EXP(GAMMALN(B\$1+C35)+GAMMALN(B\$2+B35-C35)-GAMMALN(B\$1+B\$2+B35))/E\$1	=LN(E35)
36	31	731	1		=COMBIN(B36,C36)*EXP(GAMMALN(B\$1+C36)+GAMMALN(B\$2+B36-C36)-GAMMALN(B\$1+B\$2+B36))/E\$1	=LN(E36)
37	32	326	0		=COMBIN(B37,C37)*EXP(GAMMALN(B\$1+C37)+GAMMALN(B\$2+B37-C37)-GAMMALN(B\$1+B\$2+B37))/E\$1	=LN(E37)
38	33	772	1		=COMBIN(B38,C38)*EXP(GAMMALN(B\$1+C38)+GAMMALN(B\$2+B38-C38)-GAMMALN(B\$1+B\$2+B38))/E\$1	=LN(E38)
39	34	335	1		=COMBIN(B39,C39)*EXP(GAMMALN(B\$1+C39)+GAMMALN(B\$2+B39-C39)-GAMMALN(B\$1+B\$2+B39))/E\$1	=LN(E39)
40	35	235	0		=COMBIN(B40,C40)*EXP(GAMMALN(B\$1+C40)+GAMMALN(B\$2+B40-C40)-GAMMALN(B\$1+B\$2+B40))/E\$1	=LN(E40)
41	36	218	0		=COMBIN(B41,C41)*EXP(GAMMALN(B\$1+C41)+GAMMALN(B\$2+B41-C41)-GAMMALN(B\$1+B\$2+B41))/E\$1	=LN(E41)
42	37	221	0		=COMBIN(B42,C42)*EXP(GAMMALN(B\$1+C42)+GAMMALN(B\$2+B42-C42)-GAMMALN(B\$1+B\$2+B42))/E\$1	=LN(E42)
43	38	103	1		=COMBIN(B43,C43)*EXP(GAMMALN(B\$1+C43)+GAMMALN(B\$2+B43-C43)-GAMMALN(B\$1+B\$2+B43))/E\$1	=LN(E43)
44	39	170	0		=COMBIN(B44,C44)*EXP(GAMMALN(B\$1+C44)+GAMMALN(B\$2+B44-C44)-GAMMALN(B\$1+B\$2+B44))/E\$1	=LN(E44)
45	40	45	0		=COMBIN(B45,C45)*EXP(GAMMALN(B\$1+C45)+GAMMALN(B\$2+B45-C45)-GAMMALN(B\$1+B\$2+B45))/E\$1	=LN(E45)

Problem 3 -- Model

	A	B	C	D	E	F	G	H	I
1	\alpha	0.439	B(\alpha,\beta)		0.273				
2	\beta	95.411							
3					LL =	-200.548		cutoff	0.00207
4									
5	Segment	m_s	x_s		P(X=x m)			E[p_s x_s]	Roll?
6	1	34	0		0.87448	-0.134		0.00338	Y
7	2	102	1		0.16556	-1.798		0.00727	Y
8	3	53	0		0.82334	-0.194		0.00295	Y
9	4	145	2		0.07694	-2.565		0.01013	Y
10	5	1254	62		0.00015	-8.793		0.04626	Y
11	6	144	7		0.00301	-5.805		0.03101	Y
12	7	1235	80		0.00003	-10.403		0.06044	Y
13	8	573	34		0.00014	-8.869		0.05149	Y
14	9	1083	24		0.00362	-5.622		0.02073	Y
15	10	352	5		0.03010	-3.503		0.01214	Y
16	11	817	7		0.02810	-3.572		0.00815	Y
17	12	118	0		0.70182	-0.354		0.00205	N
18	13	1049	3		0.06653	-2.710		0.00300	Y
19	14	452	3		0.06735	-2.698		0.00628	Y
20	15	338	2		0.09913	-2.311		0.00562	Y
21	16	168	0		0.63981	-0.447		0.00166	N
22	17	242	3		0.05465	-2.907		0.01018	Y
23	18	185	1		0.18091	-1.710		0.00512	Y
24	19	116	0		0.70473	-0.350		0.00207	Y
25	20	69	1		0.14588	-1.925		0.00873	Y
26	21	193	1		0.18122	-1.708		0.00498	Y
27	22	82	1		0.15531	-1.862		0.00809	Y
28	23	265	1		0.18042	-1.712		0.00399	Y
29	24	171	0		0.63664	-0.452		0.00164	N
30	25	1554	7		0.03089	-3.477		0.00451	Y
31	26	1339	4		0.05107	-2.975		0.00309	Y
32	27	1167	4		0.05197	-2.957		0.00352	Y
33	28	621	2		0.09808	-2.322		0.00340	Y
34	29	1013	1		0.13667	-1.990		0.00130	N
35	30	544	1		0.16210	-1.820		0.00225	Y
36	31	731	1		0.15052	-1.894		0.00174	N
37	32	326	0		0.52048	-0.653		0.00104	N
38	33	772	1		0.14826	-1.909		0.00166	N
39	34	335	1		0.17658	-1.734		0.00334	Y
40	35	235	0		0.57918	-0.546		0.00133	N
41	36	218	0		0.59277	-0.523		0.00140	N
42	37	221	0		0.59030	-0.527		0.00139	N
43	38	103	1		0.16596	-1.796		0.00724	Y
44	39	170	0		0.63769	-0.450		0.00165	N
45	40	45	0		0.84365	-0.170		0.00312	Y
46	41	237	0		0.57764	-0.549		0.00132	N
47	42	86	0		0.75377	-0.283		0.00241	Y
48	43	297	1		0.17887	-1.721		0.00366	Y
49	44	415	0		0.47847	-0.737		0.00086	N
50	45	187	0		0.62053	-0.477		0.00155	N
51	46	248	0		0.56944	-0.563		0.00128	N

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

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

63

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:

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

64

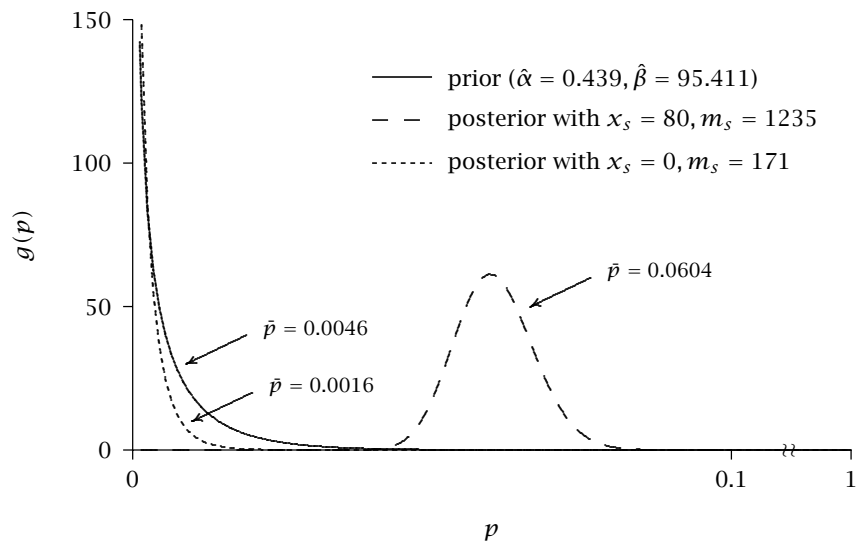
Bayes Theorem

For the beta-binomial model, we have:

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

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

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”

67

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.)

68

Problem 3 -- Model (b)

	A	B	C	D	E	F	G	H	I
1	\alpha	0.439		B(\alpha,\beta)	0.2733				
2	\beta	95.411							
3					LL = -200.548		cutoff		=(3343/10000)/161.5
4									
5	Segment	m_s	x_s		P(X=x m)			E[p_s x_s]	Roll?
6	1	34	0		0.8745	-0.1341		=(B\$1+C6)/(B\$1+B\$2+B6)	=IF(H6>=\$3,"Y","N")
7	2	102	1		0.1656	-1.7984		=(B\$1+C7)/(B\$1+B\$2+B7)	=IF(H7>=\$3,"Y","N")
8	3	53	0		0.8233	-0.1944		=(B\$1+C8)/(B\$1+B\$2+B8)	=IF(H8>=\$3,"Y","N")
9	4	145	2		0.0769	-2.5647		=(B\$1+C9)/(B\$1+B\$2+B9)	=IF(H9>=\$3,"Y","N")
10	5	1254	62		0.0002	-8.7933		=(B\$1+C10)/(B\$1+B\$2+B10)	=IF(H10>=\$3,"Y","N")
11	6	144	7		0.003	-5.8046		=(B\$1+C11)/(B\$1+B\$2+B11)	=IF(H11>=\$3,"Y","N")
12	7	1235	80		0	-10.4029		=(B\$1+C12)/(B\$1+B\$2+B12)	=IF(H12>=\$3,"Y","N")
13	8	573	34		0.0001	-8.8693		=(B\$1+C13)/(B\$1+B\$2+B13)	=IF(H13>=\$3,"Y","N")
14	9	1083	24		0.0036	-5.6216		=(B\$1+C14)/(B\$1+B\$2+B14)	=IF(H14>=\$3,"Y","N")
15	10	352	5		0.0301	-3.5032		=(B\$1+C15)/(B\$1+B\$2+B15)	=IF(H15>=\$3,"Y","N")
16	11	817	7		0.0281	-3.5719		=(B\$1+C16)/(B\$1+B\$2+B16)	=IF(H16>=\$3,"Y","N")
17	12	118	0		0.7018	-0.3541		=(B\$1+C17)/(B\$1+B\$2+B17)	=IF(H17>=\$3,"Y","N")
18	13	1049	3		0.0665	-2.7102		=(B\$1+C18)/(B\$1+B\$2+B18)	=IF(H18>=\$3,"Y","N")
19	14	452	3		0.0674	-2.6978		=(B\$1+C19)/(B\$1+B\$2+B19)	=IF(H19>=\$3,"Y","N")
20	15	338	2		0.0991	-2.3113		=(B\$1+C20)/(B\$1+B\$2+B20)	=IF(H20>=\$3,"Y","N")
21	16	168	0		0.6398	-0.4466		=(B\$1+C21)/(B\$1+B\$2+B21)	=IF(H21>=\$3,"Y","N")
22	17	242	3		0.0547	-2.9067		=(B\$1+C22)/(B\$1+B\$2+B22)	=IF(H22>=\$3,"Y","N")
23	18	185	1		0.1809	-1.7098		=(B\$1+C23)/(B\$1+B\$2+B23)	=IF(H23>=\$3,"Y","N")
24	19	116	0		0.7047	-0.3499		=(B\$1+C24)/(B\$1+B\$2+B24)	=IF(H24>=\$3,"Y","N")
25	20	69	1		0.1459	-1.925		=(B\$1+C25)/(B\$1+B\$2+B25)	=IF(H25>=\$3,"Y","N")
26	21	193	1		0.1812	-1.708		=(B\$1+C26)/(B\$1+B\$2+B26)	=IF(H26>=\$3,"Y","N")
27	22	82	1		0.1553	-1.8623		=(B\$1+C27)/(B\$1+B\$2+B27)	=IF(H27>=\$3,"Y","N")
28	23	265	1		0.1804	-1.7125		=(B\$1+C28)/(B\$1+B\$2+B28)	=IF(H28>=\$3,"Y","N")
29	24	171	0		0.6366	-0.4516		=(B\$1+C29)/(B\$1+B\$2+B29)	=IF(H29>=\$3,"Y","N")
30	25	1554	7		0.0309	-3.4774		=(B\$1+C30)/(B\$1+B\$2+B30)	=IF(H30>=\$3,"Y","N")
31	26	1339	4		0.0511	-2.9745		=(B\$1+C31)/(B\$1+B\$2+B31)	=IF(H31>=\$3,"Y","N")
32	27	1167	4		0.052	-2.9572		=(B\$1+C32)/(B\$1+B\$2+B32)	=IF(H32>=\$3,"Y","N")
33	28	621	2		0.0981	-2.3219		=(B\$1+C33)/(B\$1+B\$2+B33)	=IF(H33>=\$3,"Y","N")
34	29	1013	1		0.1367	-1.9902		=(B\$1+C34)/(B\$1+B\$2+B34)	=IF(H34>=\$3,"Y","N")
35	30	544	1		0.1621	-1.8195		=(B\$1+C35)/(B\$1+B\$2+B35)	=IF(H35>=\$3,"Y","N")
36	31	731	1		0.1505	-1.8936		=(B\$1+C36)/(B\$1+B\$2+B36)	=IF(H36>=\$3,"Y","N")

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.

69

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

70

Further Reading

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

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

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

Further Applications and Tools/ Modeling Issues

Recap

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

73

Further Applications: Timing Models

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

74

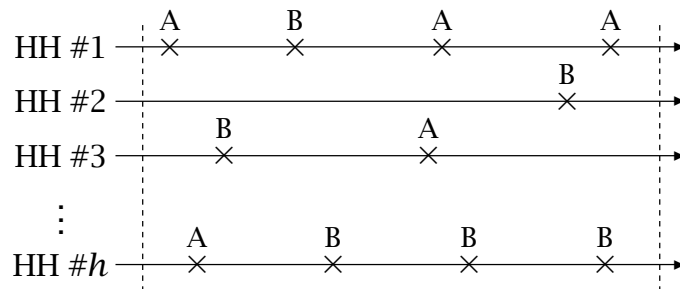
Further Applications: Count Models

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

75

Further Applications: “Choice” Models

- Brand choice



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

76

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)

77

Further Issues

Relaxing usual assumptions:

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

The basic models are quite robust to these departures.

78

Extensions

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

79

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

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

80