Problem 1:
Predicting New Product Trial
(Modeling Timing Data)
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.

<table>
<thead>
<tr>
<th>Week</th>
<th># Households</th>
<th>Week</th>
<th># Households</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>8</td>
<td>13</td>
<td>68</td>
</tr>
<tr>
<td>2</td>
<td>14</td>
<td>14</td>
<td>72</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>15</td>
<td>75</td>
</tr>
<tr>
<td>4</td>
<td>32</td>
<td>16</td>
<td>81</td>
</tr>
<tr>
<td>5</td>
<td>40</td>
<td>17</td>
<td>90</td>
</tr>
<tr>
<td>6</td>
<td>47</td>
<td>18</td>
<td>94</td>
</tr>
<tr>
<td>7</td>
<td>50</td>
<td>19</td>
<td>96</td>
</tr>
<tr>
<td>8</td>
<td>52</td>
<td>20</td>
<td>96</td>
</tr>
<tr>
<td>9</td>
<td>57</td>
<td>21</td>
<td>96</td>
</tr>
<tr>
<td>10</td>
<td>60</td>
<td>22</td>
<td>97</td>
</tr>
<tr>
<td>11</td>
<td>65</td>
<td>23</td>
<td>97</td>
</tr>
<tr>
<td>12</td>
<td>67</td>
<td>24</td>
<td>101</td>
</tr>
</tbody>
</table>
Krunchy Bits Cumulative Trial

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
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:

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

• \( \lambda \) represents the individual’s trial rate.
The Market-Level Model

Assume two segments of consumers:

<table>
<thead>
<tr>
<th>Segment</th>
<th>Description</th>
<th>Size</th>
<th>( \lambda )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>ever triers</td>
<td>( p )</td>
<td>( \theta )</td>
</tr>
<tr>
<td>2</td>
<td>never triers</td>
<td>( 1 - p )</td>
<td>0</td>
</tr>
</tbody>
</table>

\[
P(T \leq t) = P(T \leq t | \text{ever trier}) \times P(\text{ever trier}) +\]
\[
P(T \leq t | \text{never trier}) \times P(\text{never trier})
= pF(t | \lambda = \theta) + (1 - p)F(t | \lambda = 0)
= p(1 - e^{-\theta t})
\]

→ the “exponential w/ never triers” model

Estimating Model Parameters

The log-likelihood function is defined as:

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

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

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

- Let $T(t) =$ cumulative # households that have made a trial purchase by time $t$:

  $$E[T(t)] = N \times \hat{F}(t)$$

  $$= N \hat{p}(1 - e^{-\hat{\delta}t}), \quad t = 1, 2, \ldots$$

  where $N$ is the panel size.

- Use projection factors for market-level estimates.

Cumulative Trial Forecast

![Cumulative Trial Forecast Graph](image)
Extending the Basic Model

- The “exponential w/ never triers” model assumes all triers have the same underlying trial rate $\theta$ — a bit simplistic.

- Allow for multiple trier “segments” each with a different (latent) trial rate:

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

- Replace the discrete distribution with a continuous distribution.

Distribution of Trial Rates

![Distribution of Trial Rates Graph](image-url)
Distribution of Trial Rates

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

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

where \( r \) is the “shape” parameter and \( \alpha \) is the “scale” parameter.

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

Illustrative Gamma Density Functions

![Illustrative Gamma Density Functions](image-url)
**Alternative Market-Level Model**

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

$$P(T \leq t) = \int_0^\infty P(T \leq t | \lambda) \, g(\lambda) \, d\lambda$$

$$= 1 - \left( \frac{\alpha}{\alpha + t} \right)^r$$

We call this the “exponential-gamma” model.

**Estimating Model Parameters**

The log-likelihood function is defined as:

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

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

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.
Concepts and Tools Introduced

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

Further Reading


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

• Prediction
  – To settings (e.g., time periods) beyond the observation period
  – Conditional on past behavior

• Profiling behavioral propensities of individuals

• Structural analysis — basic understanding of the behavior being modeled

• 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 traits.
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 \textit{mixing distribution}.

(v) Derive the corresponding \textit{aggregate} or \textit{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: 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

Problem 2:
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)

<table>
<thead>
<tr>
<th># Exposures</th>
<th># People</th>
<th># Exposures</th>
<th># People</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>48</td>
<td>12</td>
<td>5</td>
</tr>
<tr>
<td>1</td>
<td>37</td>
<td>13</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>24</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>11</td>
<td>19</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>20</td>
<td>2</td>
</tr>
<tr>
<td>9</td>
<td>7</td>
<td>21</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>6</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>
**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).
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 $\lambda$:

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

- Exposure rates ($\lambda$) are distributed across the population according to the gamma distribution:

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

Modeling One Week Exposures

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

$$P(X = x) = \int_0^\infty P(X = x | \lambda) g(\lambda) \, d\lambda$$

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

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

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

Computing NBD Probabilities

• Note that
\[
\frac{P(X = x)}{P(X = x - 1)} = \frac{r + x - 1}{x(\alpha + 1)}
\]

• We can therefore compute NBD probabilities using the following forward recursion formula:
\[
P(X = x) = \begin{cases} 
\left( \frac{\alpha}{\alpha + 1} \right)^r & \text{x = 0} \\
\frac{r + x - 1}{x(\alpha + 1)} \times P(X = x - 1) & \text{x \geq 1}
\end{cases}
\]

Estimating Model Parameters

The log-likelihood function is defined as:
\[
LL(r, \alpha|\text{data}) = 48 \times \ln[P(X = 0)] + 37 \times \ln[P(X = 1)] + 30 \times \ln[P(X = 2)] + \ldots + 1 \times \ln[P(X = 23)]
\]

The maximum value of the log-likelihood function is \( LL = -649.7 \), which occurs at \( \hat{r} = 0.969 \) and \( \hat{\alpha} = 0.218 \).
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 \( \lambda \), then \( X(t) \) has a Poisson distribution with rate parameter \( \lambda t \):

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

NBD for a Non-Unit Time Period

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

\[
P(X(t) = x) = \int_0^\infty P(X(t) = x | \lambda) g(\lambda) \, d\lambda
\]

\[
= \frac{\Gamma(r + x)}{\Gamma(r) x!} \left( \frac{\alpha}{\alpha + t} \right)^r \left( \frac{t}{\alpha + t} \right)^x
\]

- The mean of this distribution is given by \( E[X(t)] = rt / \alpha \).
Exposure Distributions: 1 week vs. 4 week

Effectiveness of Monthly Showing

- For $t = 4$, we have:
  - $P(X(t) = 0) = 0.057$, and
  - $E[X(t)] = 17.78$

- It follows that:
  - Reach = $1 - P(X(t) = 0)$
    = 94.3%
  - Frequency = $E[X(t)]/(1 - P(X(t) = 0))$
    = 18.9
  - GRPs = $100 \times E[X(t)]$
    = 1778
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


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,

   \[ RR \text{ > } \frac{\text{cost of each mailing}}{\text{unit margin}} \]
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

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

Notation:

\[ N_s = \text{size of segment } s \ (s = 1, \ldots, S) \]
\[ m_s = \# \text{members of segment } s \text{ tested} \]
\[ X_s = \# \text{responses to test in segment } s \]

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

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

How is \( p_s \) distributed across segments?

The Beta Binomial Model

- Heterogeneity in \( p_s \) is captured using the beta distribution

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

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

- For a randomly chosen segment,

\[ P(X_s = x_s|m_s) = \int_0^1 P(X_s = x_s|m_s, p_s) \ g(p_s) \ dp_s \]

\[ = \binom{m_s}{x_s} \frac{B(\alpha + x_s, \beta + m_s - x_s)}{B(\alpha, \beta)} \]
Illustrative Beta Density Functions

![Graph showing Beta density functions with different parameters: \( \alpha = 5, \beta = 5 \), \( \alpha = 1, \beta = 1 \), \( \alpha = 0.5, \beta = 0.5 \), \( \alpha = 1.5, \beta = 0.5 \), \( \alpha = 0.5, \beta = 1.5 \), \( \alpha = 2, \beta = 4 \).]

Shape of the Beta Density

![Diagram showing the shape of the Beta density with values of \( \alpha \) and \( \beta \), indicating U-shaped, J-shaped, and reverse-J-shaped distributions with mode at \( \frac{\alpha-1}{\alpha+\beta-2} \).]
Estimating Model Parameters

The log-likelihood function is defined as:

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

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

Applying the Model

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

$$g(p_s|X_s = x_s, m_s) = \frac{P(X_s = x_s|m_s, p_s) g(p_s)}{\int_0^\infty P(X_s = x_s|m_s, p_s) g(p_s) \, dp_s}$$

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

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

Now 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

<table>
<thead>
<tr>
<th></th>
<th>Standard</th>
<th>Manager</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td># Segments (Rule)</td>
<td>51</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td># Segments (Act.)</td>
<td>46</td>
<td>71</td>
<td>53</td>
</tr>
<tr>
<td>Contacts</td>
<td>682,392</td>
<td>858,728</td>
<td>732,675</td>
</tr>
<tr>
<td>Responses</td>
<td>4,463</td>
<td>4,804</td>
<td>4,582</td>
</tr>
<tr>
<td>Profit</td>
<td>$492,651</td>
<td>$488,773</td>
<td>$495,060</td>
</tr>
</tbody>
</table>

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


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.

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 website browsing session
Further Applications: Count Models

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

Further Applications: “Choice” Models

- Brand choice

  HH #1\[\times \] A \[\times \] B \[\times \] A \[\times \] A \[\times \] B

  HH #2\[\times \] B \[\times \] A \[\times \]

  HH #3\[\times \] A \[\times \] B \[\times \]

  \vdots

  HH #h\[\times \] A \[\times \] B \[\times \] B \[\times \] B

- Media exposure
- Multibrand choice (BB $\rightarrow$ 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)

**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.
Extensions

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

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

http://brucehardie.com/talks.html