Modeling Marketing Promotion Choices

Sean Salleh
Operations and Decision Technologies
Convex Optimization Winter 2011
Final Project
Contents

- Motivation
- A Simplified Story
- Problem
- Data
- A Multinomial Logit Choice Model
- A Maximum Likelihood Model
- A Maximum Entropy Model
- Solving the Convex Program
- Model Comparison
- Conclusion
Motivation

- A realistic story

YouGo.com is a start-up company that gives discounts to consumers to attend restaurants, visual arts, live entertainment, sports, exclusive shopping, gourmet tasting, and interactive events (e.g.: dancing, painting).

- Based on a consumer’s search behavior, YouGo recommends to a consumer a set of promotions. The consumer can then pick at most one promotion from the set.

- YouGo wants to optimize their recommendation system so that a good set of promotions is shown to a consumer, that is, there is a high probability that a consumer will pick a promotion.
Motivation

- A realistic story

Live music

Comedy
A Simplified Story

- **Assumptions**
  - There is one **homogeneous** consumer segment with 21 consumers.
  - Each of the consumers is shown a promotion set comprised of 2 promotions, comedy and live music.
  - Each consumer picks exactly one promoted event.

- **Average time** to events (from the consumer’s zipcode) is the only attribute that YouGo will consider.
  - Each consumer uses the **average time** to events (shown by YouGo) in making their decision; consumers may use other factors that are unknown to YouGo in making their decision.
Problem

- **Problem**
  Using the data, estimate a *multinomial logit* choice model to predict the probability that a consumer will pick a specific event based on average time to event.

- **Goals**
  - Estimate the predictive model via **maximum likelihood**.
  - Estimate the predictive model via **maximum entropy**.
  - Compare the fit of 3 models.
  - Suggest analytic strategy to use.
Data

<table>
<thead>
<tr>
<th>Consumer</th>
<th>Choice</th>
<th>Average time (Live Music)</th>
<th>Average time (Comedy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Comedy</td>
<td>52.9</td>
<td>4.4</td>
</tr>
<tr>
<td>2</td>
<td>Comedy</td>
<td>4.1</td>
<td>28.5</td>
</tr>
<tr>
<td>3</td>
<td>Live Music</td>
<td>4.1</td>
<td>86.9</td>
</tr>
<tr>
<td>4</td>
<td>Comedy</td>
<td>56.2</td>
<td>31.6</td>
</tr>
<tr>
<td>5</td>
<td>Comedy</td>
<td>51.8</td>
<td>20.2</td>
</tr>
<tr>
<td>6</td>
<td>Live Music</td>
<td>0.2</td>
<td>91.2</td>
</tr>
<tr>
<td>7</td>
<td>Live Music</td>
<td>27.6</td>
<td>79.7</td>
</tr>
<tr>
<td>8</td>
<td>Comedy</td>
<td>89.9</td>
<td>2.2</td>
</tr>
<tr>
<td>9</td>
<td>Comedy</td>
<td>41.5</td>
<td>24.5</td>
</tr>
<tr>
<td>10</td>
<td>Comedy</td>
<td>95</td>
<td>43.5</td>
</tr>
<tr>
<td>11</td>
<td>Comedy</td>
<td>99.1</td>
<td>8.4</td>
</tr>
<tr>
<td>12</td>
<td>Live Music</td>
<td>18.5</td>
<td>84</td>
</tr>
<tr>
<td>13</td>
<td>Live Music</td>
<td>82</td>
<td>38</td>
</tr>
<tr>
<td>14</td>
<td>Comedy</td>
<td>8.6</td>
<td>1.6</td>
</tr>
<tr>
<td>15</td>
<td>Live Music</td>
<td>22.5</td>
<td>74.1</td>
</tr>
<tr>
<td>16</td>
<td>Live Music</td>
<td>51.4</td>
<td>83.8</td>
</tr>
<tr>
<td>17</td>
<td>Comedy</td>
<td>81</td>
<td>19.2</td>
</tr>
<tr>
<td>18</td>
<td>Live Music</td>
<td>51</td>
<td>85</td>
</tr>
<tr>
<td>19</td>
<td>Live Music</td>
<td>62.2</td>
<td>90.1</td>
</tr>
<tr>
<td>20</td>
<td>Comedy</td>
<td>95.1</td>
<td>22.2</td>
</tr>
<tr>
<td>21</td>
<td>Live Music</td>
<td>41.6</td>
<td>91.5</td>
</tr>
</tbody>
</table>

Time is in minutes.
A Multinomial Logit Choice Model

- A common utility maximizing choice model yields...

\[
P_n(\alpha, \beta, C) = \frac{\exp(\alpha + \beta \cdot \text{time}_i^n)}{\sum_{j \in C} \exp(\alpha + \beta \cdot \text{time}_j^n)}
\]

\[C = \{\text{live music, comedy}\}.
\]

\[\text{time}_i^n = \text{time for consumer } n \text{ for event } i.
\]

\[P_n(\alpha, \beta, C) = \text{probability consumer } n \text{ picks event } i \text{ with parameters } \alpha \text{ and } \beta \text{ and choice set } C.
\]

Model has many behavioral assumptions on consumers.
Common solution method: Newton-Raphson heuristic because of nonlinear equations from first-order conditions are difficult.
A Maximum Entropy Model

- A convex program with linear constraints

\[
\begin{align*}
\min & \sum_{n=1}^{21} \sum_{i \in C} P_n(i \mid C) \log P_n(i \mid C) \\
\text{s.t.} & \sum_{n=1}^{21} \sum_{i \in C} P_n(i \mid C) \text{time}_i^n = \sum_{n=1}^{21} y_i^n \text{time}_i^n \\
& \sum_{n=1}^{21} P_n(i \mid C) = \sum_{n=1}^{21} y_i^n, \forall i \in C \\
& \sum_{i \in C} P_n(i \mid C) = 1, \forall n \\
& P_n(i \mid C) > 0, \forall n, i
\end{align*}
\]

\[C = \{\text{live music, comedy}\}.\]

\[y_i^n = \text{choice (1 or 0) that consumer } n \text{ picked event } i.\]

\[P_n(i \mid C) = \text{probability consumer } n \text{ picks event } i.\]
A Maximum Entropy Model

- A convex program with linear constraints

We solve for the **multinomial logit** choice model parameters by solving **simpler nonlinear equations** obtained from first-order conditions from the Lagrangian.

- The **Lagrange multipliers** are the parameters...
  \[ \alpha, \beta, \]

  and the equations *naturally* give rise to...

\[
P_n(\alpha, \beta, C) = \frac{\exp(\alpha + \beta \cdot time_i^n)}{\sum_{j \in C} \exp(\alpha + \beta \cdot time_j^n)}
\]
Solving the Maximum Entropy Model

- **Software**
  - **Solver:** MINOS.
  - **Modeling language:** AMPL.

**Model file:**

```plaintext
# INITIALIZE
param numob;
param numalt;
param eps;
set Sob := 1..numob;
set Salt := 1..numalt;
param Dec(n in Sob, a in Salt); param Time(n in Sob, a in Salt);
var p(n in Sob, a in Salt) >= eps;

MINIMISE

subject to constraint1: sum(n in Sob, a in Salt) p(n,a)*log(p(n,a));

subject to constraint2(a in Salt):
  sum(n in Sob) p(n,a) = sum(n in Sob, a in Salt) Dec[n,a]*Time[n,a];

subject to constraint3(n in Sob):
  sum(a in Salt) p(n,a) = 1;
```

**Data file:**

```plaintext
param numob := 31;
param numalt := 24;
param eps := 1e-6;

# Data file for 31 movies, 24 categories

# Movie ratings data
```

---

The Paul Merage School of Business
University of California - Irvine

---

INNOVATE • COLLABORATE • GROW
Model Comparison

- Live Music

ML = Maximum likelihood, MLh = ML with heuristic, ME = maximum entropy
Model Comparison

- Comedy

ML = Maximum likelihood, MLh = ML with heuristic, ME = maximum entropy
## Model Comparison

### Statistics

<table>
<thead>
<tr>
<th></th>
<th>ML</th>
<th>MLh</th>
<th>ME</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Total deviation</strong></td>
<td>7.417</td>
<td>7.645</td>
<td>7.220</td>
</tr>
<tr>
<td><strong>Average deviation</strong></td>
<td>0.353</td>
<td>0.364</td>
<td>0.344</td>
</tr>
<tr>
<td><strong>Standard deviation</strong></td>
<td>0.493</td>
<td>0.485</td>
<td>0.483</td>
</tr>
</tbody>
</table>

Better! Well, just a bit.
Conclusion

- The maximum entropy (convex program) method is...
  - more accurate for this dataset (others were tested with similar results).
  - easier to solve, due to simpler nonlinear equations.
  - requires less assumptions about consumers ...great!

- Next step
  - Although ML and ME methods are equivalent, determine why the ME method appears to be more accurate.