Modeling Marketing Promotion Choices

Sean Salleh

Operations and Decision Technologies

Convex Optimization Winter 2011

Final Project

Contents

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

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

Time is in minutes.

A Multinomial Logit Choice Model

- **A common utility maximizing choice model yields…**

$$P_n(\alpha,\beta,C) = \frac{\exp\left(\alpha + \beta \cdot time_i^n\right)}{\sum_{j \in C} \exp\left(\alpha + \beta \cdot time_j^n\right)}$$

$C = \{\text{live music, comedy}\}.$

$time_i^n =$ time for consumer n for event i.

$P_n(\alpha,\beta,C) =$ probability consumer n picks event i with parameters α and β and choice set C.

Model has many behavioral assumptions on consumers.

A Maximum Likelihood Model

- **A concave unconstrained program**

$$\max \sum_{n=1}^{21} \sum_{i \in C} y_i^n \log P_n(\alpha, \beta, C)$$

Recall that...

$$P_n(\alpha, \beta, C) = \frac{\exp(\alpha + \beta \cdot time_i^n)}{\sum_{j \in C} \exp(\alpha + \beta \cdot time_j^n)}$$

$C =$ {live music, comedy}.

$y_i^n =$ choice (1 or 0) that consumer n picked event i.

$P_n(\alpha, \beta, C) =$ probability consumer n picks event i with parameters α and β and choice set C.

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**

$$\min \sum_{n=1}^{21} \sum_{i \in C} P_n(i \mid C) \log P_n(i \mid C)$$

$$\text{s.t. } \sum_{n=1}^{21} \sum_{i \in C} P_n(i \mid C) time_i^n = \sum_{n=1}^{21} \sum_{i \in C} y_i^n time_i^n$$

$$\sum_{n=1}^{21} P_n(i \mid C) = \sum_{n=1}^{21} y_i^n, \forall i \in C$$

$$\sum_{i \in C} P_n(i \mid C) = 1, \forall n$$

$$P_n(i \mid C) > 0, \forall n, i$$

$C = \{\text{live music, comedy}\}$.

$y_i^n =$ choice (1 or 0) that consumer n picked event i.

$P_n(i \mid C) =$ probability consumer n 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…

$$P_n(\alpha, \beta, C) = \frac{\exp\left(\alpha + \beta \cdot time_i^n\right)}{\sum_{j \in C} \exp\left(\alpha + \beta \cdot time_j^n\right)}$$

Solving the Maximum Entropy Model

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

Model file:

```
#INITIALIZATION

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;


#OBJECTIVE

minimize objective:
        sum{n in Sob, a in Salt} p[n,a]*log(p[n,a]);


#CONSTRAINTS

subject to constraint1:
        sum{n in Sob, a in Salt} p[n,a]*time[n,a] = sum{n in Sob, a in Salt} dec[n,a]*time[n,a];

subject to constraint2{a in Salt}:
        sum{n in Sob} p[n,a] = sum{n in Sob} dec[n,a];

subject to constraint3{n in Sob}:
        sum{a in Salt} p[n,a] = 1;
```

Data file:

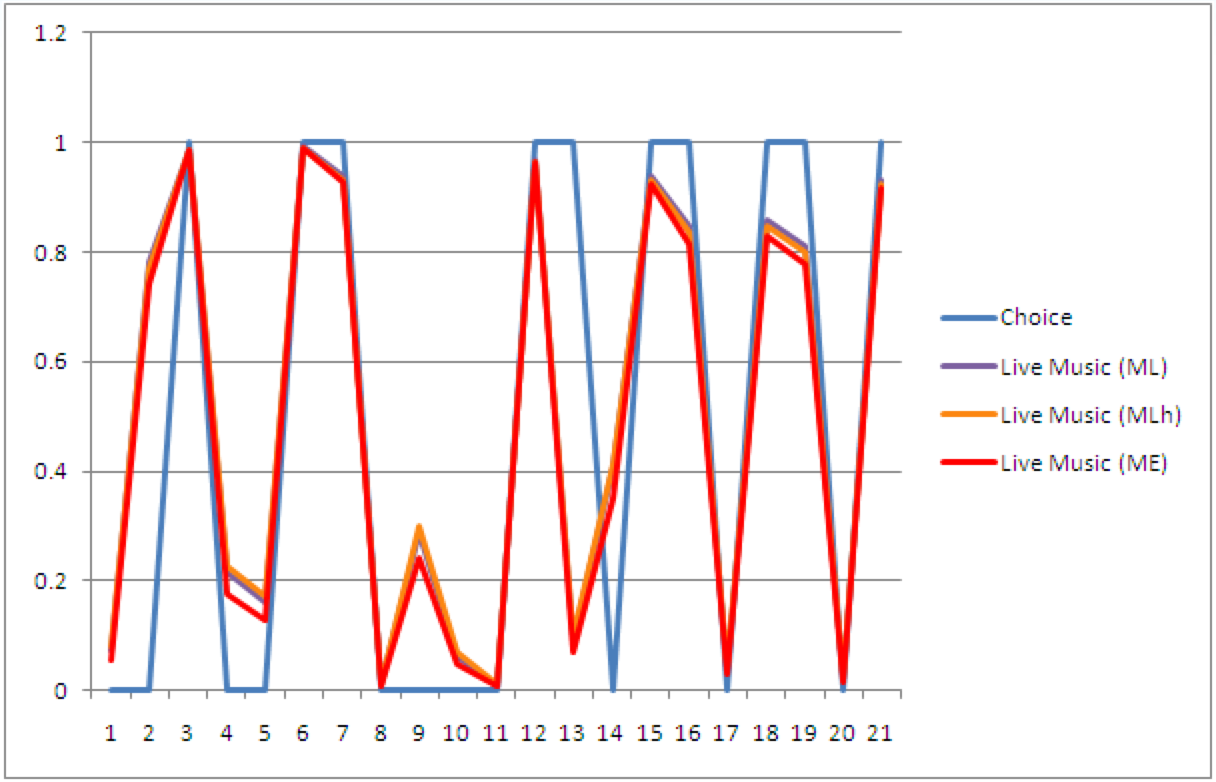
```
param numob := 21;
param numalt := 2;
param eps := 1e-6;

#Obs binary decisions, with live music = 1 and comedy = 2.
param dec : 1 2 :=
1       0       1
2       0       1
3       1       0
4       0       1
5       0       1
6       1       0
7       1       0
8       0       1
9       0       1
10      0       1
11      0       1
12      1       0
13      1       0
14      0       1
15      1       0
16      1       0
17      0       1
18      1       0
19      1       0
20      0       1
21      1       0;


#Times for live music = 1 and comedy = 2.
param time : 1 2 :=
1       52.9    4.4
2       4.1     28.5
3       4.1     86.9
4       56.2    31.6
5       51.8    20.2
6       0.2     91.2
7       27.6    79.7
8       89.9    2.2
9       41.5    24.5
10      95      43.5
11      99.1    8.4
12      18.5    84
13      82      38
14      8.6     1.6
15      22.5    74.1
16      51.4    83.8
17      81      19.2
18      51      85
19      62.2    90.1
20      95.1    22.2
21      41.6    91.5;
```

Model Comparison

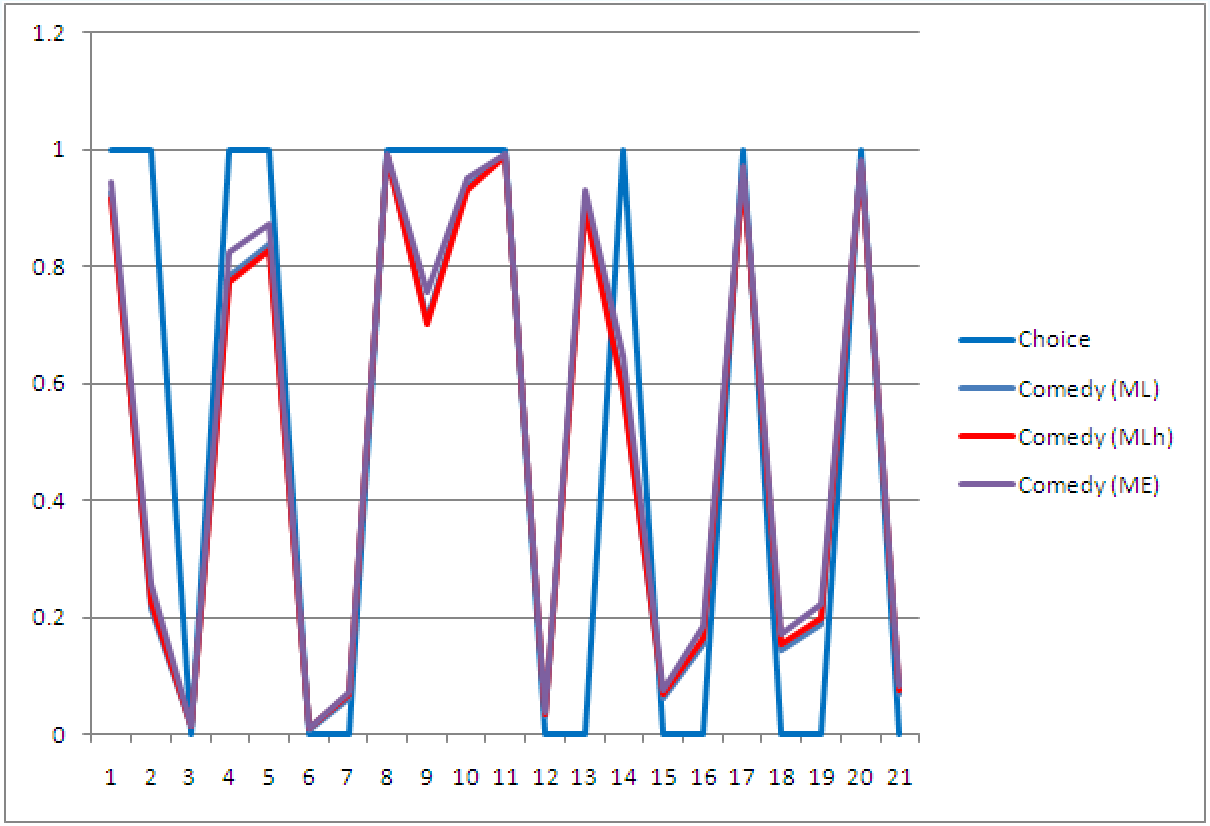
- **Live Music**



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**

	ML	MLh	ME
Total deviation	7.417	7.645	7.220
Average deviation	0.353	0.364	0.344
Standard deviation	0.493	0.485	0.483

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.