Discrete choice and discrete optimization: a continuous quest for integration

Michel Bierlaire

Transport and Mobility Laboratory
School of Architecture, Civil and Environmental Engineering
Ecole Polytechnique Fédérale de Lausanne

December 18, 2015

Outline

Optimization problem

Given...

the demand

Find...

the best configuration of the system.

Example: airline

Context

- An airline considers to propose various destinations $i = \{1, \ldots, J\}$ to its customers.
- Each potential destination i is served by an aircraft, with capacity c_i.
- The price of the ticket for destination i is p_i.
- The demand is known: W_i passengers want to travel to i.
- The fixed cost of operating a flight to destination i is F_i.
- The airline cannot invest more than a budget B.

Question

What destinations should the airline serve to maximize its revenues?

Example: airline

Decisions variables

$y_i \in \{0, 1\}$: 1 if destination i is served, 0 otherwise.

Maximize revenues

$$\max \sum_{i=1}^{J} \min(W_i, c_i) p_i y_i$$

Constraints

$$\sum_{i=1}^{J} F_i y_i \leq B$$

TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

Example: airline

Integer linear optimization problem

- Decision variables are integers.
- Objective function and constraints are linear.
- Here: knapsack problem.

Solving the problem

- Branch and bound
- Cutting planes

Example: airline

Pricing

- What price p_i should the airline propose?

$$\max \sum_{i=1}^{J} \min(W_i, c_i) p_i y_i$$

Issues

- Non linear objective
- Unbounded problem

Example: airline

Unbounded problem

- As demand is constant, the airline can make money with very high prices.
- We need to take into account the impact of price on demand.

Logit model

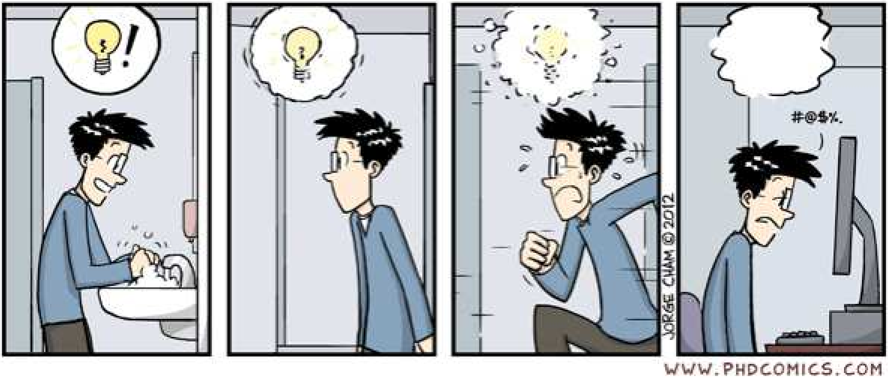
$$W_i = \sum_n P_n(i|p_i, z_{in}, s_n)$$

$$P_n(i|p_i, z_{in}, s_n) = \frac{y_i e^{V_{in}(p_i, z_{in}, s_n)}}{\sum_{j \in \mathcal{C}} y_j e^{V_{jn}(p_j, z_{jn}, s_n)}}.$$

The problem becomes highly non linear.

Outline

The main idea



TRANSP-OR

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

The main idea

Linearization

Hopeless to linearize the logit formula (we tried...)

First principles

Each customer solves an optimization problem

Solution

Use the utility and not the probability

A linear formulation

Utility function

$$U_{in} = V_{in} + \varepsilon_{in} = \sum_k \beta_k x_{ink} + f(z_{in}) + \varepsilon_{in}.$$

Simulation

- Assume a distribution for ε_{in}
- E.g. logit: i.i.d. extreme value
- Draw R realizations ξ_{inr}, $r = 1, \ldots, R$
- The choice problem becomes deterministic

Scenarios

Draws

- Draw R realizations ξ_{inr}, $r = 1, \ldots, R$
- We obtain R scenarios

$$U_{inr} = \sum_k \beta_k x_{ink} + f(z_{in}) + \xi_{inr}.$$

- For each scenario r, we can identify the largest utility.
- It corresponds to the chosen alternative.

Comparing utilities

Variables

$$\mu_{ijnr} = \begin{cases} 1 & \text{if } U_{inr} \geq U_{jnr}, \\ 0 & \text{if } U_{inr} < U_{jnr}. \end{cases}$$

Constraints

$$(\mu_{ijnr} - 1)M_{nr} \leq U_{inr} - U_{jnr} \leq \mu_{ijnr}M_{nr}, \forall i, j, n, r.$$

where

$$|U_{inr} - U_{jnr}| \leq M_{nr}, \forall i, j,$$

Comparing utilities

$$(\mu_{ijnr}-1)M_{nr} \leq U_{inr}-U_{jnr} \leq \mu_{ijnr}M_{nr}, \forall i,j,n,r.$$

Constraints: $\mu_{ijnr}=1$

$$0 \leq U_{inr}-U_{jnr} \leq M_{nr}, \forall i,j,n,r.$$

$$U_{jnr} \leq U_{inr}, \forall i,j,n,r.$$

Constraints: $\mu_{ijnr}=0$

$$-M_{nr} \leq U_{inr}-U_{jnr} \leq 0, \forall i,j,n,r.$$

$$U_{inr} \leq U_{jnr}, \forall i,j,n,r.$$

Comparing utilities

$$(\mu_{ijnr} - 1)M_{nr} \leq U_{inr} - U_{jnr} \leq \mu_{ijnr} M_{nr}, \forall i, j, n, r.$$

Equivalence if no tie

$$\begin{aligned}
\mu_{ijnr} = 1 &\Longrightarrow U_{inr} \geq U_{jnr} \\
\mu_{ijnr} = 0 &\Longrightarrow U_{inr} \leq U_{jnr} \\
U_{inr} > U_{jnr} &\Longrightarrow \mu_{ijnr} = 1 \\
U_{inr} < U_{jnr} &\Longrightarrow \mu_{ijnr} = 0
\end{aligned}$$

Accounting for availabilities

Motivation

- If $y_i = 0$, alternative i is not available.
- Its utility should not be involved in any constraint.

New variables: two alternatives are both available

$$\eta_{ij} = y_i y_j$$

Linearization:

$$y_i + y_j \leq 1 + \eta_{ij},$$
$$\eta_{ij} \leq y_i,$$
$$\eta_{ij} \leq y_j.$$

Comparing utilities of available alternatives

Constraints

$$M_{nr}\eta_{ij} - 2M_{nr} \leq U_{inr} - U_{jnr} - M_{nr}\mu_{ijnr} \leq (1-\eta_{ij})M_{nr}, \forall i,j,n,r.$$

$\eta_{ij} = 1$ and $\mu_{ijnr} = 1$

$$0 \leq U_{inr} - U_{jnr} \leq M_{nr}, \forall i,j,n,r.$$

$\eta_{ij} = 1$ and $\mu_{ijnr} = 0$

$$-M_{nr} \leq U_{inr} - U_{jnr} \leq 0, \forall i,j,n,r.$$

Comparing utilities of available alternatives

Constraints

$$M_{nr}\eta_{ij} - 2M_{nr} \leq U_{inr} - U_{jnr} - M_{nr}\mu_{ijnr} \leq (1-\eta_{ij})M_{nr}, \forall i,j,n,r.$$

$\eta_{ij} = 0$ and $\mu_{ijnr} = 1$

$$-M_{nr} \leq U_{inr} - U_{jnr} \leq 2M_{nr}, \forall i,j,n,r,$$

$\eta_{ij} = 0$ and $\mu_{ijnr} = 0$

$$-2M_{nr} \leq U_{inr} - U_{jnr} \leq M_{nr}, \forall i,j,n,r,$$

Comparing utilities of available alternatives

Valid inequalities

$$\mu_{ijnr} \leq y_i, \qquad \forall i, j, n, r,$$
$$\mu_{ijnr} + \mu_{jinr} \leq 1, \qquad \forall i, j, n, r.$$

The choice

Variables

$$w_{inr} = \begin{cases} 1 & \text{if } n \text{ chooses } i \text{ in scenario } r, \\ 0 & \text{otherwise} \end{cases}$$

Maximum utility

$$w_{inr} \leq \mu_{ijnr}, \forall i, j, n, r.$$

Availability

$$w_{inr} \leq y_i, \forall i, n, r.$$

The choice

One choice

$$\sum_{i \in \mathcal{C}} w_{inr} = 1, \forall n, r.$$

Demand and revenues

Demand

$$W_i = \frac{1}{R} \sum_{n=1}^{n} \sum_{r=1}^{R} w_{inr}.$$

Revenues

$$R_i = \frac{1}{R} \sum_{n=1}^{N} p_i \sum_{r=1}^{R} w_{inr}.$$

Outline

A simple example

Data

- $\mathcal{C}$: set of movies
- Population of N individuals
- Utility function:
  $U_{in} = \beta_{in} p_{in} + f(z_{in}) + \varepsilon_{in}$

Decision variables

- What movies to propose? y_i
- What price? p_{in}

Demand model

Logit model

Probability that n chooses movie i:

$$P(i|y, p_n, z_n) = \frac{y_i e^{\beta_{in} p_{in} + f(z_{in})}}{\sum_j y_j e^{\beta_{jn} p_{jn} + f(z_{jn})}}$$

Total revenue:

$$\sum_{i \in C} y_i \sum_{n=1}^{N} p_{in} P(i|y, p_n, z_n)$$

Non linear and non convex in the decision variables

Example: programming movie theaters

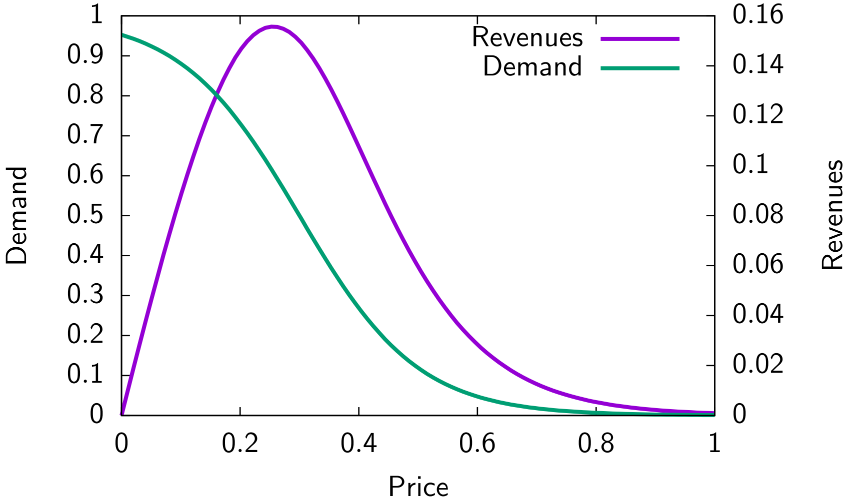
Data

- Two alternatives: my theater (m) and the competition (c)
- We assume an homogeneous population of N individuals

$$U_c = 0 + \varepsilon_c$$
$$U_m = \beta_c p_m + \varepsilon_m$$

- $\beta_c < 0$
- Logit model: ε_m i.i.d. EV

Demand and revenues



Optimization (with GLPK)

Data

- $N = 1$
- $R = 100$
- $U_m = -10p_m + 3$
- Prices: 0.10, 0.20, 0.30, 0.40, 0.50

Results

- Optimum price: 0.3
- Demand: 56%
- Revenues: 0.168

Heterogeneous population

Two groups in the population

$$U_{in} = \beta_n p_i + c_n$$

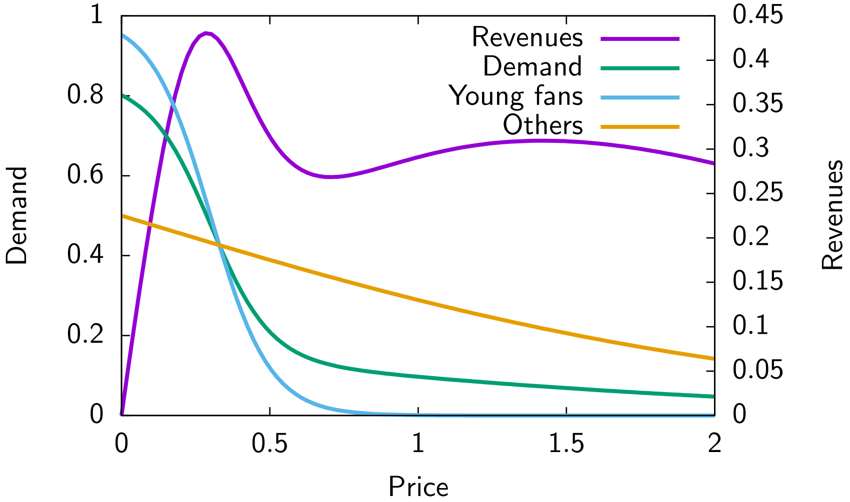
Young fans: 2/3

$\beta_1 = -10,\ c_1 = 3$

Others: 1/3

$\beta_1 = -0.9,\ c_1 = 0$

Demand and revenues



Optimization

Data

- $N = 3$
- $R = 100$
- $U_{m1} = -10p_m + 3$
- $U_{m2} = -0.9p_m$
- Prices: 0.3, 0.7, 1.1, 1.5, 1.9

Results

- Optimum price: 0.3
- Customer 1 (fan): 60% [theory: 50 %]
- Customer 2 (fan) : 49% [theory: 50 %]
- Customer 3 (other) : 45% [theory: 43 %]
- Demand: 1.54 (51%)
- Revenues: 0.48

Two theaters, different types of films

Two theaters, different types of films

Theater m

- Expensive
- Star Wars Episode VII

Theater k

- Cheap
- Tinker Tailor Soldier Spy

Heterogeneous demand

- Two third of the population is young (price sensitive)
- One third of the population is old (less price sensitive)

Two theaters, different types of films

Data

- Theaters m and k
- $N = 6$
- $R = 10$
- $U_{mn} = -10p_m + \textcircled{4}$, $n = 1, 2, 4, 5$
- $U_{mn} = -0.9p_m$, $n = 3, 6$
- $U_{kn} = -10p_k + \textcircled{0}$, $n = 1, 2, 4, 5$
- $U_{kn} = -0.9p_k$, $n = 3, 6$
- Prices m: 1.0, 1.2, 1.4, 1.6, 1.8
- Prices k: half price

Theater m

- Optimum price m: 1.6
- 4 young customers: 0
- 2 old customers: 0.5
- Demand: 0.5 (8.3%)
- Revenues: 0.8

Theater k

- Optimum price m: 0.5
- Young customers: 0.8
- Old customers: 1.5
- Demand: 2.3 (38%)
- Revenues: 1.15

Two theaters, same type of films

Theater m

- Expensive
- Star Wars Episode VII

Theater k

- Cheap
- Star Wars Episode VIII

Heterogeneous demand

- Two third of the population is young (price sensitive)
- One third of the population is old (less price sensitive)

Two theaters, same type of films

Data

- Theaters m and k
- $N = 6$
- $R = 10$
- $U_{mn} = -10p_m + \textcircled{4}$, $n = 1, 2, 4, 5$
- $U_{mn} = -0.9p_m$, $n = 3, 6$
- $U_{kn} = -10p_k + \textcircled{4}$, $n = 1, 2, 4, 5$
- $U_{kn} = -0.9p_k$, $n = 3, 6$
- Prices m: 1.0, 1.2, 1.4, 1.6, 1.8
- Prices k: half price

Theater m

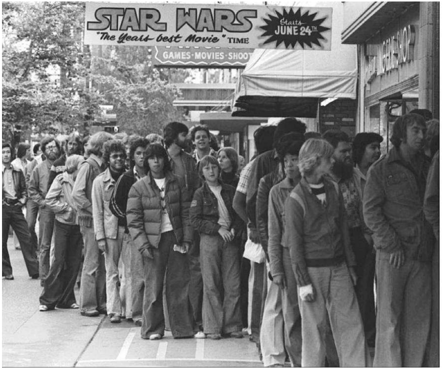
- Optimum price m: 1.8
- Young customers: 0
- Old customers: 1.9
- Demand: 1.9 (31.7%)
- Revenues: 3.42

Theater k

Closed

Extension: dealing with capacities

- Demand may exceed supply
- Not every choice can be accommodated
- Difficulty: who has access?
- Assumption: priority list is exogenous



Outline

Summary

Discrete choice in optimization

- Non linear and non convex
- Idea: use utility instead of probability
- Rely on simulation to capture stochasticity
- Not restricted to logit
- Sole assumption: endogenous variable appears linearly in the utility

Proposed formulation

- General: not designed for a specific application or context.
- Flexible: wide variety of demand and supply models.
- Scalable: the level of complexity can be adjusted.
- Integrated: not sequential.
- Operational: can be solved efficiently.

Dynamic extension

Price and capacity vary over time

Ongoing research

Revenue management

Airlines, train operators, etc.

Decomposition methods

- Scenarios are (almost) independent from each other (except objective function)
- Individuals are also loosely coupled (except for capacity constraints)

Thank you!

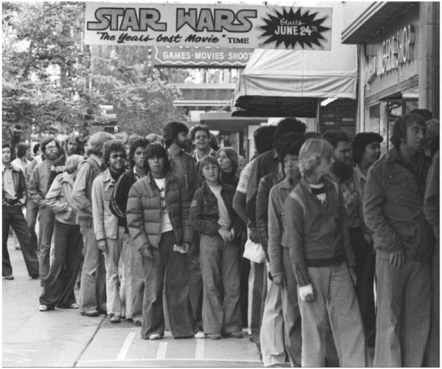
Questions?

Outline

1. Introduction
2. Integrated framework
3. A simple example
   - A linear formulation
   - Example: one theater
   - Example: two theaters
4. Summary
5. Appendix: dealing with capacities
   - Example: two theaters

Dealing with capacities

- Demand may exceed supply
- Not every choice can be accommodated
- Difficulty: who has access?
- Assumption: priority list is exogenous



Priority list

Application dependent

- First in, first out
- Frequent travelers
- Subscribers
- ...

In this framework

The list of customers must be sorted

Dealing with capacities

Variables

- y_{in}: decision of the operator
- y_{inr}: availability

Constraints

$$\sum_{n=1}^{N} w_{inr} \leq c_i$$

$$y_{inr} \leq y_{in}$$

$$y_{i(n+1)r} \leq y_{inr}$$

Constraints

$$c_i(1 - y_{inr}) \leq \sum_{m=1}^{n-1} w_{imr} + (1 - y_{in})c_{\max}$$

$y_{in} = 1,\ y_{inr} = 1$

$$0 \leq \sum_{m=1}^{n-1} w_{imr}$$

$y_{in} = 1,\ y_{inr} = 0$

$$c_i \leq \sum_{m=1}^{n-1} w_{imr}$$

$y_{in} = 0,\ y_{inr} = 0$

$$c_i \leq \sum_{m=1}^{n-1} w_{imr} + c_{\max}$$

Constraints

$$\sum_{m=1}^{n-1} w_{imr} + (1 - y_{in})c_{\max} \leq (c_i - 1)y_{inr} + \max(n, c_{\max})(1 - y_{inr})$$

$y_{in} = 1, \; y_{inr} = 1$

$$1 + \sum_{m=1}^{n-1} w_{imr} \leq c_i$$

$y_{in} = 1, \; y_{inr} = 0$

$$\sum_{m=1}^{n-1} w_{imr} \leq \max(n, c_{\max})$$

$y_{in} = 0, \; y_{inr} = 0$

$$\sum_{m=1}^{n-1} w_{imr} + c_{\max} \leq \max(n, c_{\max})$$

Two theaters, different types of films

Data

- Theaters m and k
- Capacity: 2
- $N = 6$
- $R = 5$
- $U_{mn} = -10p_m + 4, \; n = 1, 2, 4, 5$
- $U_{mn} = -0.9p_m, \; n = 3, 6$
- $U_{kn} = -10p_k + 0, \; n = 1, 2, 4, 5$
- $U_{kn} = -0.9p_k, \; n = 3, 6$
- Prices m: 1.0, 1.2, 1.4, 1.6, 1.8
- Prices k: half price

Theater m

- Optimum price m: 1.8
- Demand: 0.2 (3.3%)
- Revenues: 0.36

Theater k

- Optimum price m: 0.5
- Demand: 2 (33.3%)
- Revenues: 1.15

Example of two scenarios

Customer	Choice	Capacity m	Capacity k
1	0	2	2
2	0	2	2
3	k	2	1
4	0	2	1
5	0	2	1
6	k	2	0

Customer	Choice	Capacity m	Capacity k
1	0	2	2
2	k	2	1
3	0	2	1
4	k	2	0
5	0	2	0
6	0	2	0

Two theaters: all prices divided by 2

Data

- Theaters m and k
- Capacity: 2
- $N = 6$
- $R = 5$
- $U_{mn} = -10p_m + 4, \; n = 1, 2, 4, 5$
- $U_{mn} = -0.9p_m, \; n = 3, 6$
- $U_{kn} = -10p_k + 0, \; n = 1, 2, 4, 5$
- $U_{kn} = -0.9p_k, \; n = 3, 6$
- Prices m: 0.5, 0.6, 0.7, 0.8, 0.9
- Prices k: half price

Theater m

- Optimum price m: 0.5
- Demand: 1.4
- Revenues: 0.7

Theater k

- Optimum price m: 0.45
- Demand: 1.6
- Revenues: 0.72

Example of two scenarios

Customer	Choice	Capacity m	Capacity k
1	0	2	2
2	0	2	2
3	0	2	2
4	k	2	1
5	k	2	0
6	0	2	0

Customer	Choice	Capacity m	Capacity k
1	k	2	1
2	k	2	0
3	0	2	0
4	m	1	0
5	0	1	0
6	m	0	0