Columbia Business School

Automatic selection of rank-based choice models

Garrett J. van Ryzin, *Columbia University*

Gustavo Vulcano, *New York University*

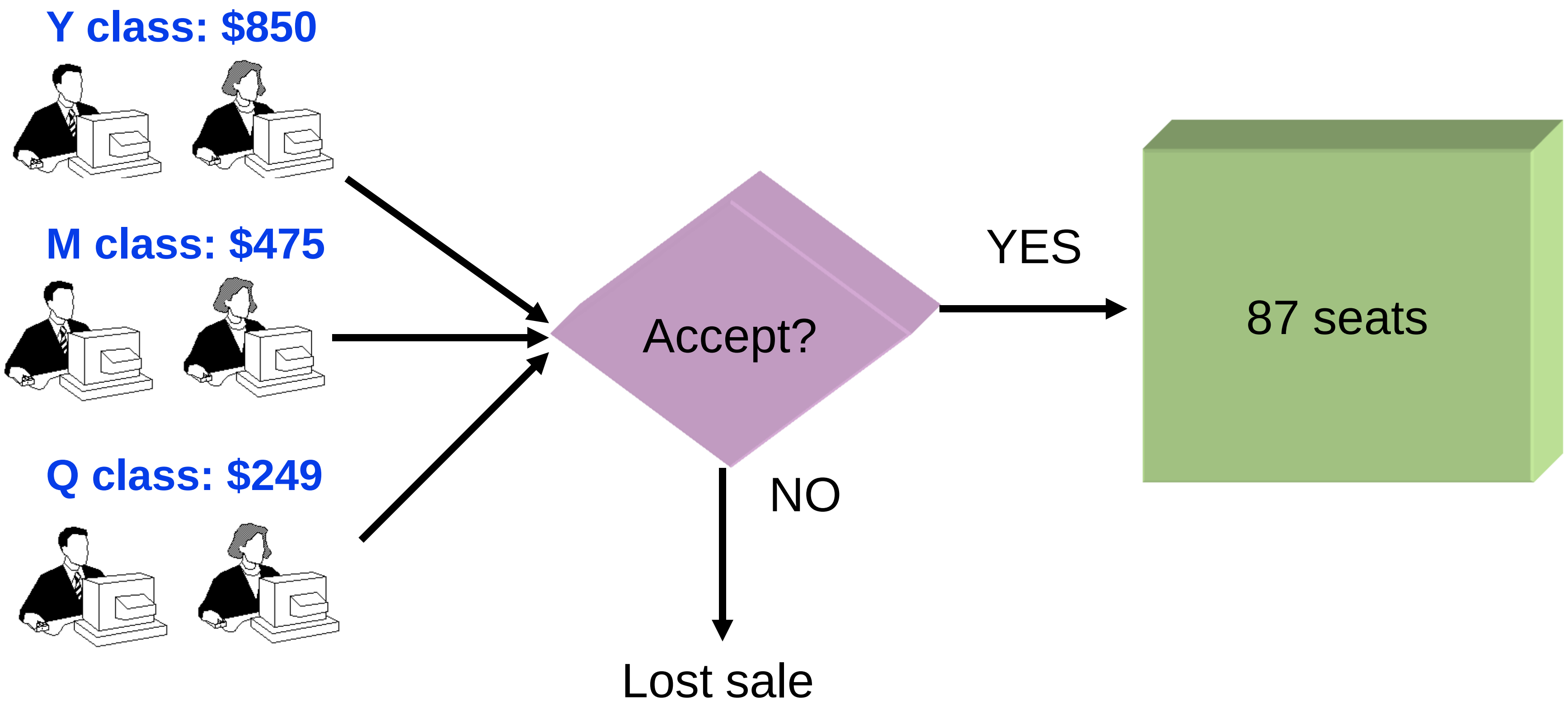
Outline

I. Rank-based models of choice in revenue management
II. Estimation problem
   – Efficient EM method for solving the MLE
   – Market discovery mechanism for automatically generating rankings
III. Numerical examples
IV. Conclusions

Traditional revenue management model

Demand is defined at the product-level…

"independent demand model"



We decide which product requests to accept or reject

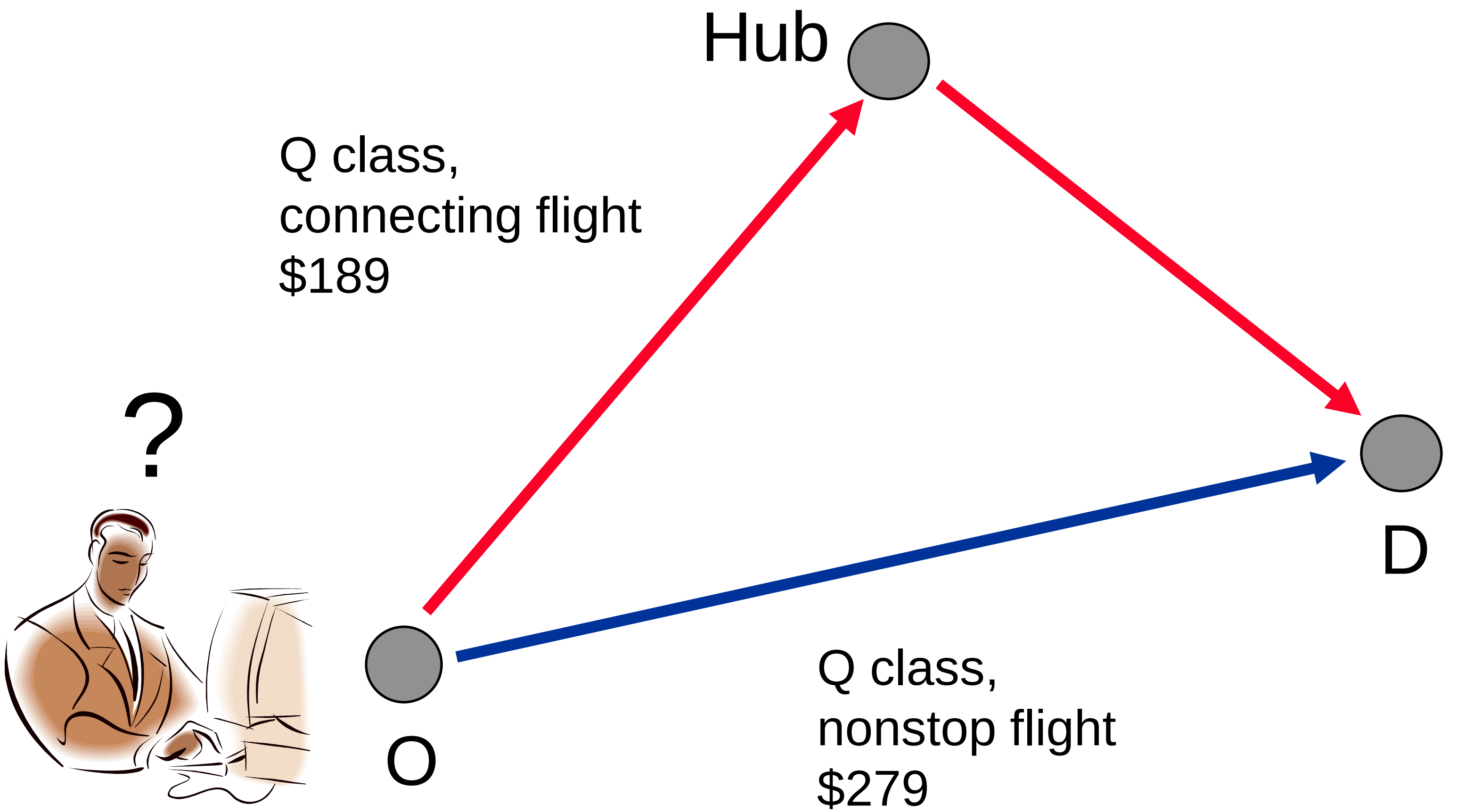
But demand for products is really the outcome of a consumer choice decision …

Which airline, route, dep. time and fare class to buy?

Find flights by: Airline / Stops Price	Delta Air Lines	US Airways	Continental Airlines	American Airlines	United Airlines	Virgin America	Alaska Airlines
Non-stop			$1,388 total $1,409		$1,388 total $1,409		
1 stop	$338 total $380	$338 total $387	$370 total $418	$430 total $472	$493 total $542	$872 total $914	$1,421 total $1,463
2+ stops						$872 total $918	

Buy now or wait?

Results in choice of paths on a network



Ex: A network model with rank-based consumer choice (van Ryzin & Vulcano 2008)

- Sequence of T customers arrive, $t=1,\ldots,T$
- Each customer t has rank preferences
  1. Ordered list of preferred products l_t

     e.g. [6,3,0,0,0,0…]

     1st choice = Product 6

     2nd choice = Product 3

     No-purchase if 1st and 2nd choice not available

  2. Quantity desired q_t

- Sample path: $\omega = \{(l_t, q_t): \ t = 1, \ldots, T\}$
- Assume parametric class of “booking limit” policies

Use *continuous acceptance* assumptions to develop a recursion for the network revenue…

Revenue-to-go from period t onward

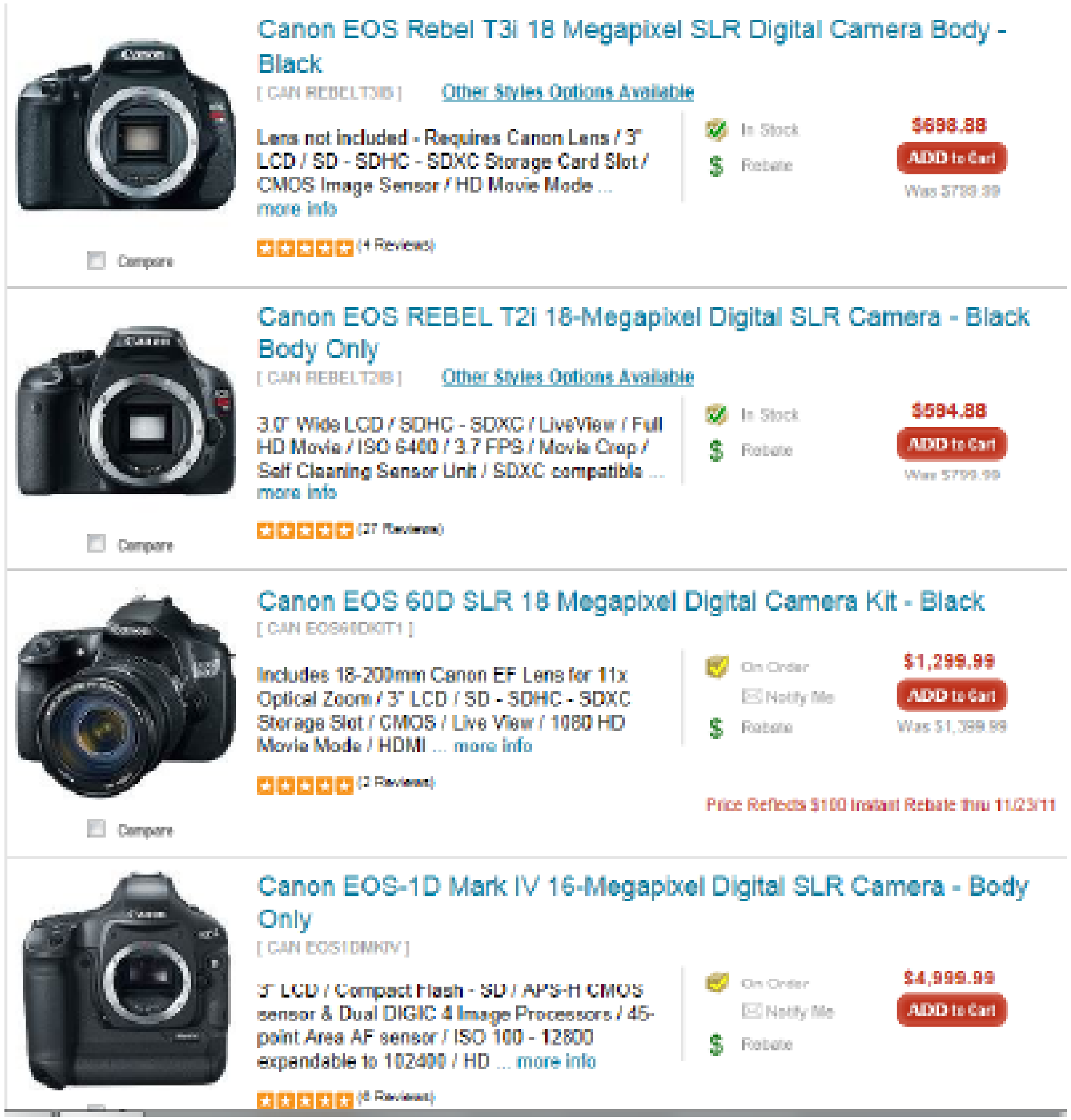
$$R_t(x(t), y, \omega) = r^T u(x(t), y, l_t, q_t) + R_{t+1}(x(t+1), y, \omega)$$

$$x(t+1) = x(t) - A^T u(x(t), y, l_t, q_t)$$

where…

$u_j(x(t), y, l, q) =$ Quantities of products purchased (vector) by customer with preference list l and desired quantity q facing remaining capacities x and protection levels y.

Other applications: Offer set optimization



Related literature in operations

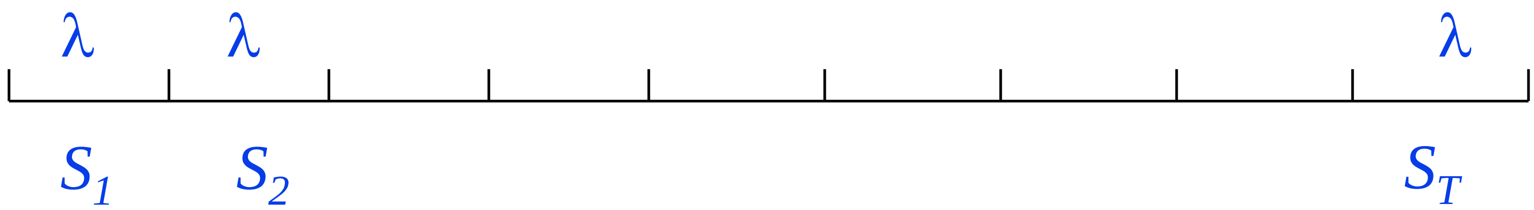
- Anupindi, Dada & Gupta (1998)
- Ja, Rao & Chandler (2001)
- Mahajan & van Ryzin (2001)
- Talluri & van Ryzin (2004)
- Conlon & Mortimer (2007)
- Kok & Fisher (2007)
- Ratliff, Rao, Narayan & Yellepeddi (2007)
- Vulcano, van Ryzin & Chaar (2007)
- Vulcano, van Ryzin & Ratliff (2010)
- Farias et al. (2013)
- Haensel and Koole (2011)

Typical availability and sales data…

What are customer underlying preferences for colors?

How many customers arrived in each week?

Demand model



Bernoulli (0,1) arrivals with probability of arrival λ in each period t

$S_t \subseteq N$ set of available products in period t

available data

$z_{it} =$ indicator of sale of product i in period t

Rank-based preferences

Mahajan & van Ryzin 2001, van Ryzin & Vulcano 2008, Farias et al. 2010

Customer types defined by <u>preference orderings</u>

σ_1 σ_2 σ_3 σ_4 σ_5 σ_6 σ_7

$P(\sigma_1)$ $P(\sigma_2)$ $P(\sigma_3)$ $P(\sigma_4)$ $P(\sigma_5)$ $P(\sigma_6)$ $P(\sigma_7)$

Probability mass function (pmf) on types

Estimation problem given known set of rankings

Given data….

$$\{z_{it} : i = 1, \ldots n, t = 1, \ldots, T\}$$

z_{it} = indicator of sales for product i in period t

$$\{S_t : t = 1, \ldots, T\}$$

Set of available products in each period

$$\{\sigma_i : i = 1, \ldots, N\}$$

Set of rankings (customer types)

Jointly estimate…..

$$\lambda$$

Arrival probability

$$x_i = P(\sigma_i) \qquad i = 1, \ldots, N$$

pmf of rank-based preference

Direct MLE approach

- Let $\mathcal{P}$ and $\bar{\mathcal{P}}$ be the set of periods with purchases and no purchases, respectively.
- The incomplete data log-likelihood function is given by

$$\mathcal{L}_I(\mathbf{x}, \lambda) = \sum_{t \in \mathcal{P}} \left(\log \lambda + \log \left(\sum_{i \in \mathcal{M}_t(j_t, S_t)} x_i \right) \right) + \sum_{\substack{t \in \bar{\mathcal{P}}, \\ \mathcal{M}_t(0, S_t) \neq \emptyset}} \log \left(\lambda \sum_{i \in \mathcal{M}_t(0, S_t)} x_i + (1 - \lambda) \right) + \sum_{\substack{t \in \bar{\mathcal{P}}, \\ \mathcal{M}_t(0, S_t) = \emptyset}} \log(1 - \lambda).$$

- The MLE estimation problem can be posed as:

$$\begin{aligned} \max_{\mathbf{x}, \lambda} \quad & \mathcal{L}_I(\mathbf{x}, \lambda) \\ \text{s.t.:} \quad & \sum_{i=1}^{N} x_i = 1, \\ & \mathbf{x} \geq 0, \\ & 0 \leq \lambda \leq 1. \end{aligned}$$

- Remarks:
  - Difficult non-linear optimization problem
  - The objective function is data-dependent and not even quasiconcave in general

Suppose we could observe each customer's type (ranking) directly

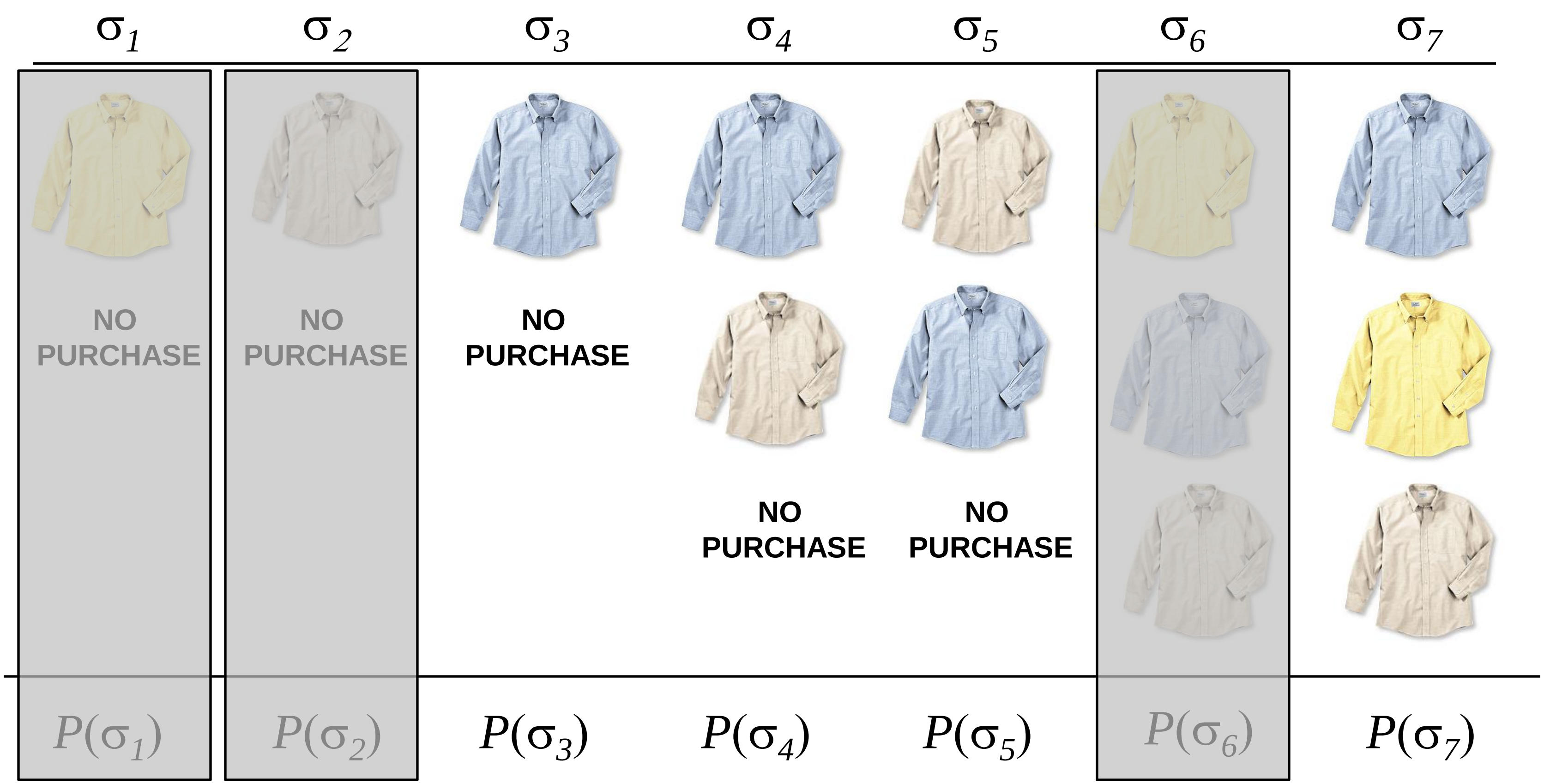
Leads to a trivial MLE problem for PMF with solution

$$x_i^* = \frac{m_i}{\sum_{k=1}^{N} m_k}$$

where $m_i =$ total number of type i customers observed

While this level of information is unrealistic, we can treat observed sales data as giving us partial information on a each customer's type.

Example: Arriving customer in period *t* bought blue when only blue and yellow were available.



$$P(\sigma_3 \,|\, \text{sales in } t) = \frac{P(\sigma_3)}{P(\sigma_3) + P(\sigma_4) + P(\sigma_5) + P(\sigma_7)}$$

Outline of EM procedure (Dempster et al. 1977)

1. Initialize PMF with an arbitrary guess (e.g. uniform)
2. Using incumbent PMF, compute conditional expected number of types given observed sales data

$$m_i = \sum_t P(\sigma_i \mid \text{sales in period } t)$$

3. Form complete-data log-likelihood and maximize to get new PMF

$$x_i^* = \frac{m_i}{\sum_{k=1}^{N} m_k}$$

4. GOTO Step 2

NOTE: Simple procedure to update estimate of λ in each step as well

EM algorithm for estimating the distribution of a set of customer types

[Input data]: Number of customer types N, availability information represented by sets S_t and transaction data $j_t, t = 1, \ldots, T$.

[Initialization]: Set $x_i := 1/N$, and $\lambda = 0.5$. Based on the availability information and transaction data, build the sets $\mathcal{M}_t(j_t, S_t) := \{i : \sigma^{(i)}(j_t) < \sigma^{(i)}(k), \forall k \in S_t, k \neq j_t\}$. Set $a_t := 0$, $t = 1, \ldots, T$.

Repeat

Set $m_i := 0, x_{it} := 0, i = 1, \ldots, N, t = 1, \ldots, T$.

[E-step]:

For $t := 1, \ldots, T$ do

[Update probabilities of customer types]

For $i \in \mathcal{M}_t(j_t, S_t)$ do

Set $x_{it} := x_i / (\sum_{h \in \mathcal{M}_t(j_t, S_t)} x_h)$.

Endfor

[Update estimates a_t]

If $j_t > 0$

Set $a_t := 1$,

Else (i.e., $j_t = 0$)

If $\mathcal{M}_t(0, S_t) = \emptyset$

Set $a_t := 0$,

Else

Set $a_t := \lambda \sum_{i \in \mathcal{M}_t(0,S_t)} x_i / (\lambda \sum_{i \in \mathcal{M}_t(0,S_t)} x_i + (1 - \lambda))$

Endif

Endif

Endfor

[Compute estimates for m_i]

For $i := 1$ to N do

For $t := 1$ to T do

Set $m_i := m_i + a_t x_{it}$

Endfor

Endfor

[M-step]:

For $i := 1$ to N do

Set $x_i := m_i / \sum_{t=1}^{T} a_t$.

Endfor

Set $\lambda := \sum_{t=1}^{T} a_t / T$.

Until Stopping criteria is met.

Complete algorithm in pseudo code

Main Results

- Convergence proof (Wu 1983)
  - All limit points of the algorithm are stationary points of the original, incomplete-data log-likelihood function

- Computational time significantly faster than direct MLE

What about the exponential number of possible customer types?

Some initial observations

- Identification is <u>not possible</u> in general
- Imagine perfect data: Every choice set $S_t \subseteq N$ is observed infinitely often.

$O(2^N)$ << $O(N!)$

#offer sets #rankings

The practical issue, then, is can one find some candidate set of rankings that explains the observed data well?

Market discovery mechanism

Given an initial set of types $\{\sigma_1,\ldots,\sigma_N\}$, our EM procedure solves…

- Let $y_t = \sum_{i \in \mathcal{M}_t(j_t, S_t)} x_i$ be the aggregated likelihood of the customer types that could pick alternative j_t in period t.
- Let $A \in \mathcal{R}^{T \times N}$ with elements $a_{ti} = 1$ if $i \in \mathcal{M}_t(j_t, S_t)$ (here, j_t could also be zero), and $a_{ti} = 0$ otherwise.
- Consider the problem:

$$
\begin{aligned}
\max_{\mathbf{y},\lambda} \quad & \mathcal{L}_I(\mathbf{y},\lambda) \\
\text{s.t.:} \quad 1 - \sum_{i=1}^{N} x_i \quad & \geq \quad 0, \qquad \text{(Dual vble.: } \beta^* \geq 0\text{)} \\
A\mathbf{x} - \mathbf{y} \quad & \geq \quad 0, \qquad \text{(Dual vbles.: } \boldsymbol{\mu}^* \geq 0\text{)} \\
1 - \lambda \quad & \geq \quad 0, \\
\mathbf{x}, \mathbf{y}, \lambda \quad & \geq \quad 0.
\end{aligned}
$$

Can we find a new "type" that improves the likelihood?

COLUMN GENERATION

Suppose that we could find a new, augmented point $(\mathbf{x}, x_{N+1})$, with corresponding column for the matrix A, $A^{(N+1)}$, such that

$$\sum_{\substack{t:\sigma^{(N+1)}(j_t)<\sigma^{(N+1)}(i), \\ \forall i \in S_t, i \neq j_t}} \mu_t^* > \beta^*. \tag{1}$$

Then, customer type $(N+1)$ defines a direction of improvement for the associated Lagrangean.

Interpretation: We are seeking a customer type for which the marginal increase in the likelihood, $\mathcal{L}_I(\mathbf{x}, \lambda)$, from shifting probability mass to the new type more than compensates for the marginal reduction, β^*, in likelihood due to taking away that probability mass from the current set of types.

Complexity of finding new types

Observations:

- Determining such a customer type $(N+1)$ can be posed as an optimization problem: Given sets S_t and coefficients μ_t^*, $1 \leq t \leq T$, we are interested in maximizing the LHS in (1).
- Related to the *linear ordering problem* on single machine, with $n+1$ jobs of unit length, but here costs are at the bundle level.

Proposition: NP-Hardness

The optimization problem

$$\max \sum_{t=1}^{T} \mu_t^* w_t$$

$$\text{s.t.:}$$

$$w_t \leq \mathbb{I}\{\sigma(j_t) < \sigma(i), \quad \forall i \in S_t, i \neq j_t\}, \quad t = 1, \ldots, T,$$

is NP-Hard.

Remarks:

- Reduction from *maximum independent set*.
- There is not even a fully polynomial-time approximation scheme unless P=NP.

A MIP to find new types

Define $O(n^2 + T)$ binary variables:

- Linear ordering: $x_{ji} = 1$ if product j goes before i in the preference list, and $x_{ji} = 0$ otherwise.
- Indicators: $w_t = 1$ if the dual variable u_t^* must be added to the sum in (1), and $w_t = 0$ otherwise.

MIP formulation

$$
\begin{array}{lrcl}
 & \max \sum_{t=1}^{T} u_t^* \, w_t & & \\
\text{s.t.:} & x_{ji} + x_{ij} & = & 1, \ \forall j, i, \ j < i, \\
 & x_{ji} + x_{il} + x_{lj} & \leq & 2, \ \forall j, i, l, \ j \neq i \neq l, \\
 & \sum_{j=1}^{n} x_{j0} & \geq & 1, \\
 & w_t & \leq & x_{j_t, i}, \quad \forall i \in S_t, j_t \neq i, \ \text{and } \forall t, \\
 & x_{ji}, w_t \in \{0, 1\}, & & 0 \leq j < i \leq n, \ 1 \leq t \leq T.
\end{array}
$$

The position of product j, $\sigma^{(N+1)}(j)$, can be determined from:

$$\sigma^{(N+1)}(j) = \sum_{i \in \mathcal{N}, i \neq j} x_{ij} + 1.$$

Implementation

Iterative estimation-discovery procedure

1. Start from a parsimonious set of types $\boldsymbol{\sigma} = \{\sigma^{(1)}, \ldots, \sigma^{(N)}\}$ (e.g., $\sigma^{(j_t)} = (j_t)$ for each of the transactions j_t we observe in the market, leading to $N \leq n$ singleton types).
2. Apply EM and get initial proportions $x_i, 1 \leq i \leq N$.
3. Eliminate types with very low probabilities (e.g., lower than $1e-4$).
4. Adjust remaining proportions so that $\sum_{i=1}^{N'} x_i = 1$, with $N' \leq N$.
5. Apply market discovery procedure using MIP formulation.
6. Add the new type to the set $\boldsymbol{\sigma}$ and go to Step 2.

Numerical example

- Start from independent demand assumption.
- MIP was solved using the MATLAB connector from IBM ILOG CPLEX v12.3.
- Discovery path followed by 5 random instances:

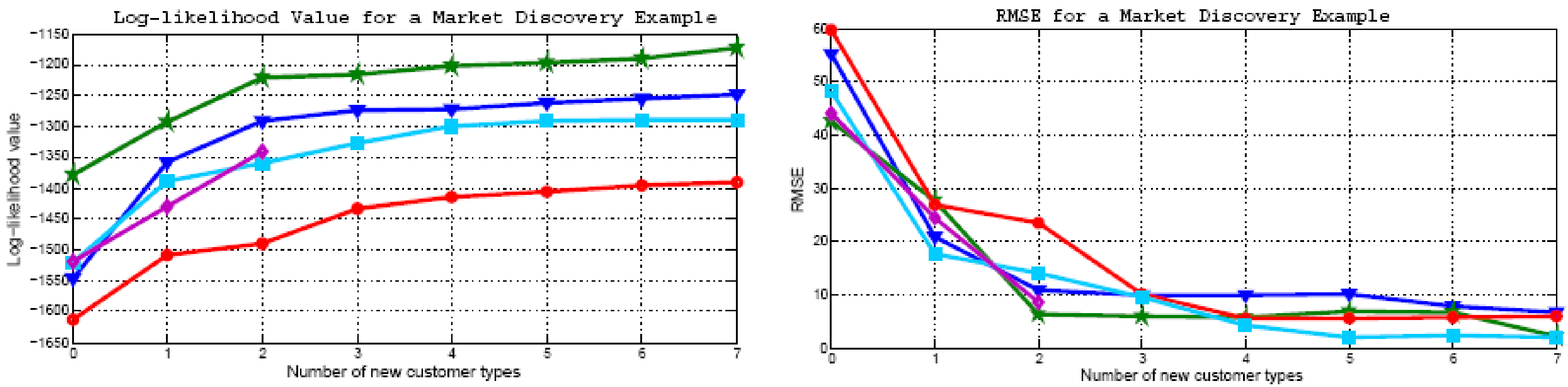


Figure 1: An example based on synthetic data generated for $n = 10$ products, $N = 5$ customer types, $T = 1000$ periods, and $\lambda = 0.7$ arrivals per period. Each product is available in each period with probability 0.5. The initial customer types are defined by the independent demand assumption built upon the observed transactions.

- Observations:
  - Dramatic improvement in the performance when adding few additional types to the independent demand assumption.
  - The beneficial effect of new types tends to be decreasing as the bank enlarges.

Additional numerical experiments

- Start from independent demand assumption.
- Baseline for comparison: True underlying demand model, and indep. demand model.
- Results based on 30 instances per experiment:

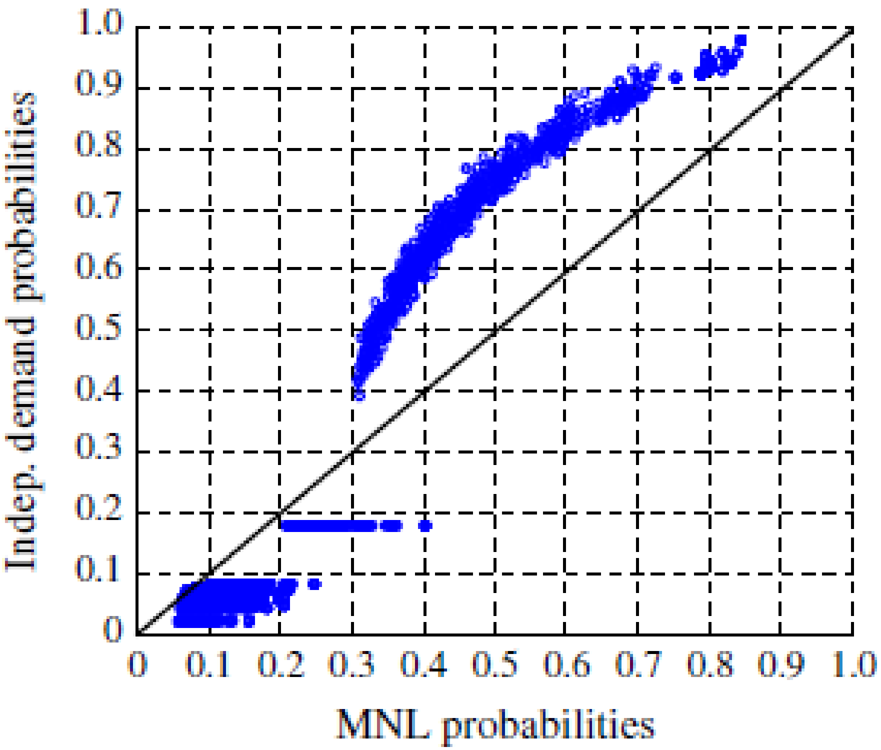
Table 7: Effect of the application of the market discovery mechanism ($n = 10, T = 1000$). Comparative results with respect to the true underlying demand model and the independent demand model. Results are 95% CIs for the mean difference.

λ	N	Log-value difference				RMSE difference			
		True demand model		Independent demand model		True demand model		Independent demand model	
		Center	Width	Center	Width	Center	Width	Center	Width
0.2	5	11.42%	10.75%	17.21%	10.07%	-61.08%	9.80%	38.43%	78.19%
	10	5.83%	6.37%	8.78%	6.18%	-65.41%	5.65%	41.33%	23.72%
	15	12.74%	10.59%	15.02%	10.32%	-60.17%	12.04%	37.37%	33.79%
0.5	5	4.08%	6.48%	14.04%	5.89%	-58.32%	8.59%	-78.80%	5.33%
	10	24.17%	15.23%	29.13%	14.24%	-72.45%	8.53%	-86.04%	3.56%
	15	44.35%	17.71%	46.52%	17.02%	-83.43%	6.15%	-88.65%	4.17%
0.8	5	7.13%	9.04%	25.76%	7.43%	-47.23%	9.22%	-94.07%	1.10%
	10	10.63%	10.84%	23.05%	9.41%	-53.78%	9.09%	-94.69%	0.97%
	15	23.85%	15.29%	32.54%	13.56%	-60.55%	10.30%	-95.97%	0.93%

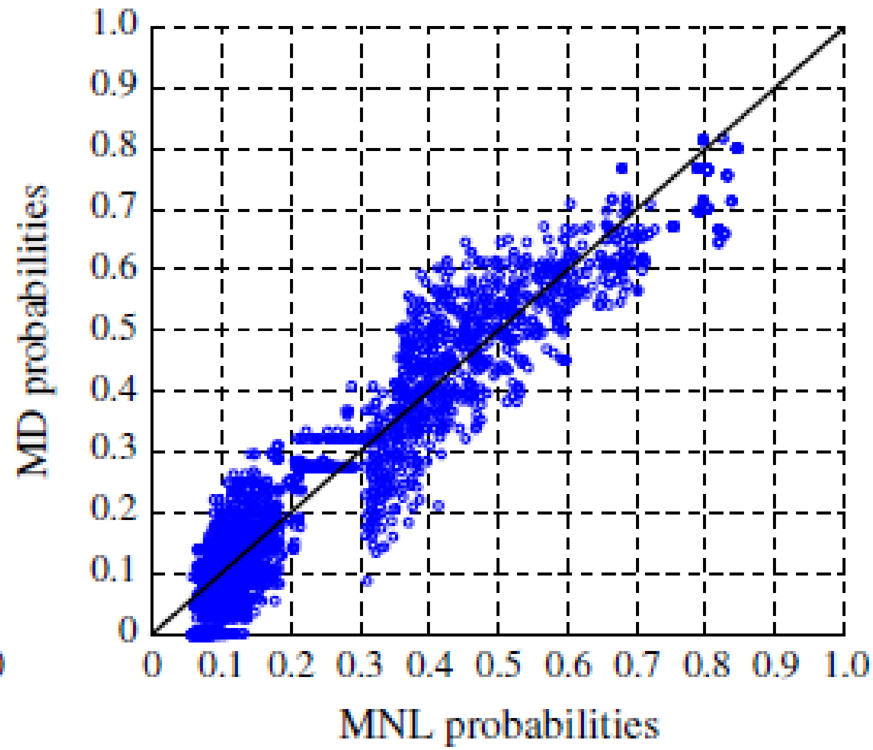
- Observations:
  - Our procedure matches (and even beats) the performance of the true underlying demand model.
  - The indep. demand model can be more appropriate for low demand cases.
  - Solving the MIP is computationally intensive, and may take around 10 seconds for each instance in the last row of the table.

Amazon DVD data: MNL model
(Farias et al. 2011)

Predicted vs. ground-truth probabilities of purchase (1000 periods)



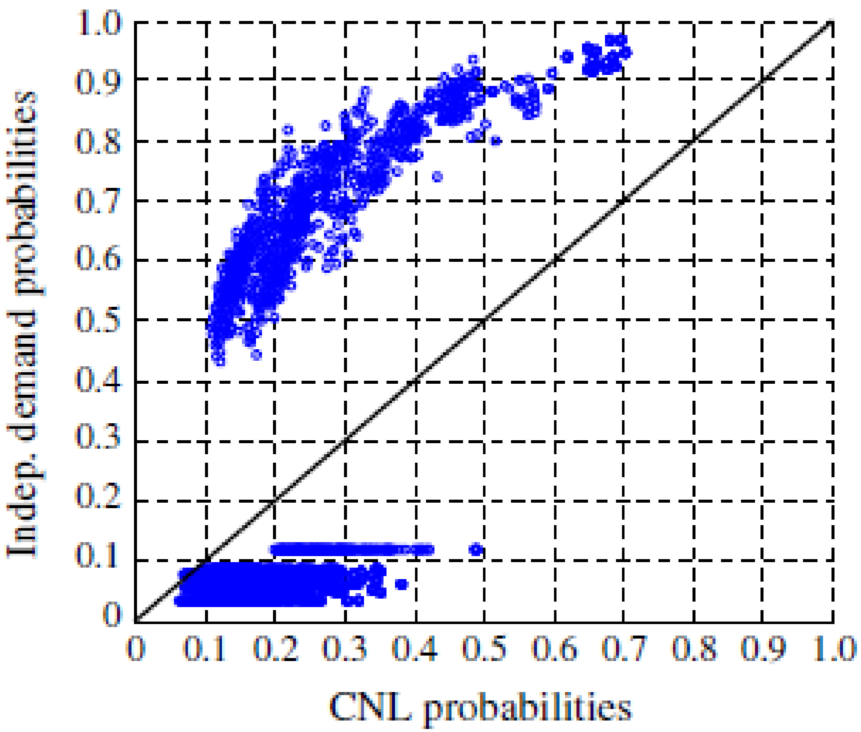
Independent Demand Model

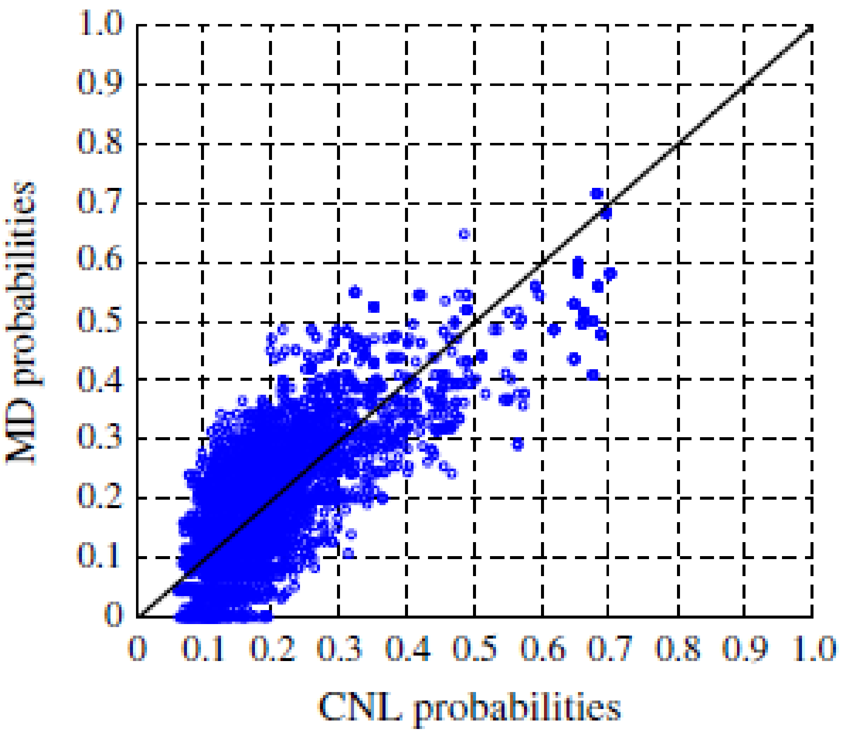


MD Demand Model

Amazon DVD data: Nested logit model

Predicted vs. ground-truth probabilities of purchase (1000 periods)



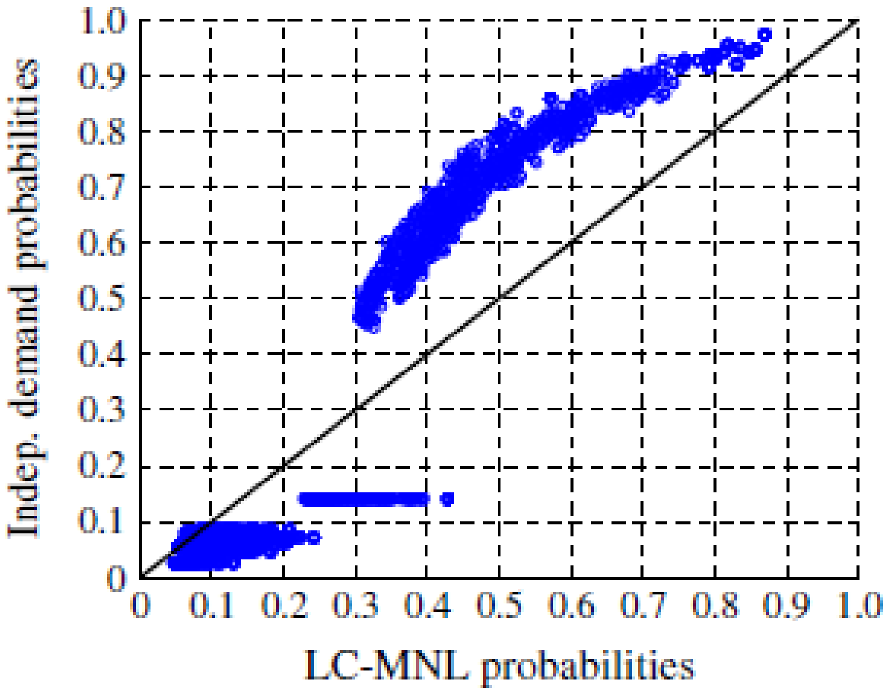


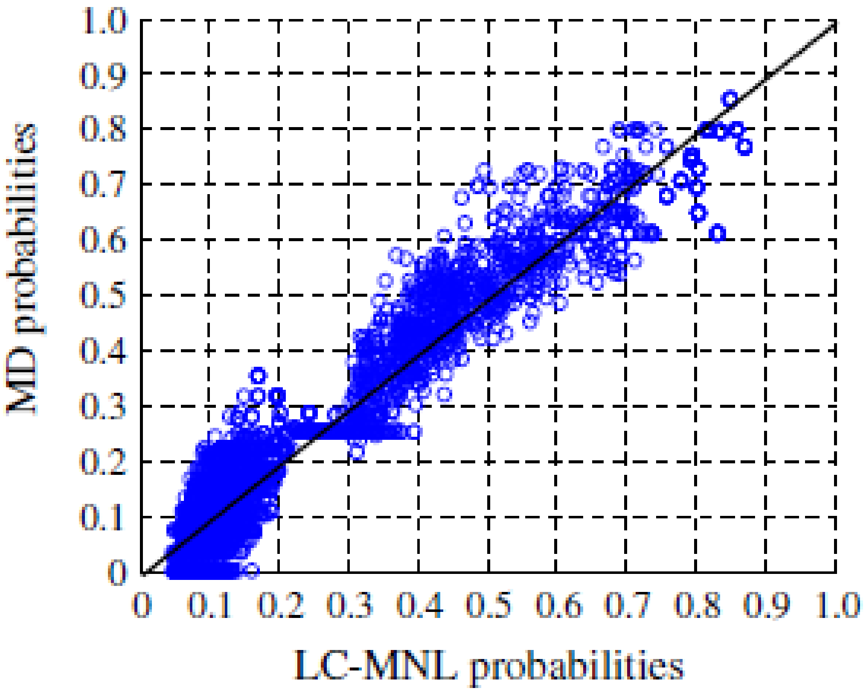
Independent Demand Model

MD Demand Model

Amazon DVD data: Latent class mixed logit model

Predicted vs. ground-truth probabilities of purchase (1000 periods)





Independent Demand Model

MD Demand Model

Real-world airline data set

- Start from independent demand assumption, and apply market discovery procedure (max.: 50 types).
- MIP was solved using the MATLAB connector from IBM ILOG CPLEX v12.3.
- Goal: Compare with *Indep* and *Single class MNL*
- 2 parallel flights (7 & 8 fare classes available, respectively):

Table 8: Estimation results for the Airline Market Example

Consideration set	Measure	Estimation method		
		Indep.	Market disc.	MNL
Flight 1	$\mathcal{L}_I(v, \lambda)$	-652.99	-564.30	-
	RMSE	1.41	5.57	2.18
Flight 2	$\mathcal{L}_I(v, \lambda)$	-734.49	-605.21	-
	RMSE	4.43	4.21	2.00
Joint	$\mathcal{L}_I(v, \lambda)$	-891.15	-717.85	-
	RMSE	6.37	4.17	6.31

- Observations:
  - Dramatic improvement of the log-likelihood value with respect to Indep.
  - RMSE is worse w.r.t. MNL under separate flights (due to very sparse data), but better for the joint flight case.
  - In general, after 5-10 types we obtained most of the improvement in log-likelihood values.
  - Final market composition consisted of at most one singleton, with a total of 37 types for the *joint flight* case.

Real-world hotel data set

- Example based on Bodea et al.(MSOM 2009), with a market share of 20% for each hotel.
- Definitions:
  - Product: Room type (e.g., king non-smoking, queen smoking, etc)
  - Period: (booking date, check-in date) pair
- Start from independent demand assumption, and apply market discovery procedure (max.: 50 types).
- MIP was solved using the MATLAB connector from IBM ILOG CPLEX v12.3.
- Goal: Compare with *Indep*

Table 9: Estimation results for the Hotel Example

Feature	Hotel 4	Hotel 5
Number of products n	9	8
Number of original bookings	373	310
Number of bookings after preprocessing	288	245
Number of periods T	1440	1225
Initial number of singletons under Indep. Demand	9	8
Indep. Demand RMSE	1.34	1.40
Indep. Demand Log-value	-977.48	-1092.46
Final number of types under Market Discovery	21, with 1 singleton	22, with 2 singletons
Market Discovery RMSE	1.08	0.89
Market Discovery Log-value	-948.83	-1062.05

- Observations:
  - Clear improvement of both the log-likelihood value and the RMSE with respect to Indep.
  - Removal of most singletons.
  - Low λ, but here products are available 75% of the time, and therefore the booking behavior is better explained by a richer demand model.

Summary

- Practical problem
  - Important revenue management applications
  - Based on realistic sales and availability data
- Simple, efficient EM procedure to estimate PMF of a known collection of types
- Market discovery mechanism based on MIP to automatically generate a “good” set of types
- Additional research questions
  - Regularization/over-fitting?
  - Fast type-generation for very large sets of alternatives