

Optimal Targeting of Customers for a Last-Minute Sale

R. Cominetti, J. Correa, Th. Rothvoss, J. San Martín

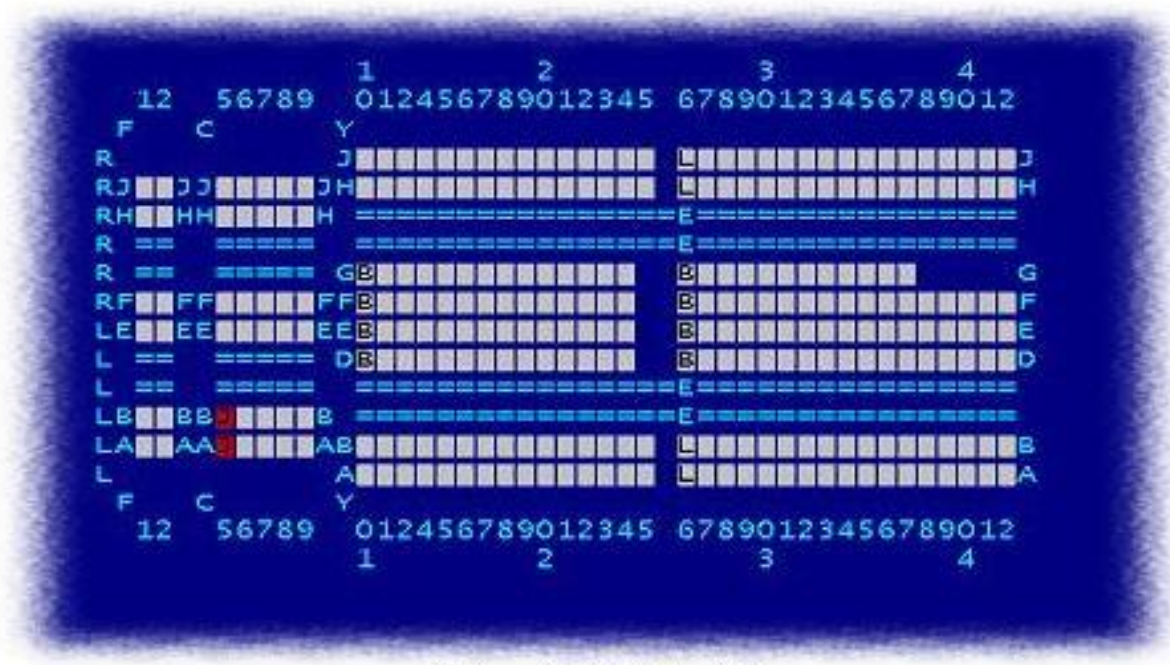


UNIVERSIDAD DE CHILE



Want to sell business class upgrades

Who should get the offer?



Airbus A340-300 Seat Map

The Process — (1) Set of customers C



(2) Address the offer to $S \subseteq C$



(3) Respondents $R \subseteq S \subseteq C$



(4) Winner chosen at random



The Single Item case

| Setting | Implication |
|---|---------------------------|
| different revenue for each client v_i | want high revenue clients |
| different acceptance probability p_i | want high pbb clients |
| last minute | no time for regret |
| until sold out | first respondent wins |

The Single Item case

| Setting | Implication |
|---|---------------------------|
| different revenue for each client v_i | want high revenue clients |
| different acceptance probability p_i | want high pbb clients |
| last minute | no time for regret |
| until sold out | first respondent wins |

Goal

balance revenues and probabilities
to maximize expected revenue

The Problem — Discrete Version

Expected revenue for $S \subseteq C$ is:

$$V_S = \mathbb{E} \left[\frac{v(R)}{|R|} \right] = \sum_{R \subseteq S} \underbrace{\frac{v(R)}{|R|}}_{\text{REVENUE IF } R \text{ ACCEPTS}} \cdot \underbrace{\prod_{i \in R} p_i}_{\text{PROB THAT } R \text{ ACCEPTS}} \cdot \underbrace{\prod_{i \in S \setminus R} (1 - p_i)}_{\text{PROB THAT } S \setminus R \text{ REJECTS}}$$

Problem: find $V^* = \max_{S \subseteq C} V_S$

The Problem — Continuous Version

Offer is made to each client with probability $x_i \in [0, 1]$

$$\mathbb{P}[i \text{ accepts}] = \mathbb{P}[Y_i = 1] = x_i \cdot p_i = y_i$$

$$V(y) = \sum_{R \subseteq C} \frac{v(R)}{|R|} \cdot \prod_{i \in R} y_i \cdot \prod_{i \notin R} (1 - y_i)$$

Problem: find $V^* = \max_{0 \leq y_i \leq p_i} V(y)$

Remarks

- $V(y)$ is affine in each variable y_i
⇒ discrete and continuous problems are equivalent

Remarks

- $V(y)$ is affine in each variable y_i
 \Rightarrow discrete and continuous problems are equivalent
- Efficient computation of objective function

$$V(y) = \mathbb{E} \left[\frac{v(R)}{|R|} \right] = \mathbb{E} \left[\frac{\sum_{i \in C} v_i Y_i}{\sum_{i \in C} Y_i} \right] = \sum_{i \in C} v_i y_i \mathbb{E} \left[\frac{1}{1 + S_i} \right]$$

distribution of $S_i = \sum_{j \neq i} Y_j$ computed by convolution

$\Rightarrow V(y)$ computed in $O(n^2)$

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?
- If so, find a threshold value V and offer to all clients with $v_i \geq V$. Assuming $v_1 \geq v_2 \geq \dots \geq v_n$, optimal threshold found in $O(n^3)$

$$\max_{1 \leq i \leq n} V_{\{1, \dots, i\}}$$

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?
- If so, find a threshold value V and offer to all clients with $v_i \geq V$. Assuming $v_1 \geq v_2 \geq \dots \geq v_n$, optimal threshold found in $O(n^4)$

Random instances: 10 customers, 200 repetitions

| Algorithm | % opt | Min ratio | Avg ratio | Time |
|----------------|-------|-----------|-----------|--------|
| <i>optimal</i> | 100.0 | 1.0000 | 1.0000 | 1611.2 |
| THRESHOLD | 93.0 | 0.9916 | 0.9998 | 23.5 |

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?
- If so, find a threshold value V and offer to all clients with $v_i \geq V$. Assuming $v_1 \geq v_2 \geq \dots \geq v_n$, optimal threshold found in $O(n^4)$
- Threshold typical in *revenue management*... but not always optimal!

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?
- If so, find a threshold value V and offer to all clients with $v_i \geq V$. Assuming $v_1 \geq v_2 \geq \dots \geq v_n$, optimal threshold found in $O(n^4)$
- Threshold typical in *revenue management*... but not always optimal!
- Sorting by probability or expected value is also sub-optimal

Threshold Strategies

- If offer is made to a customer reporting v_i , shouldn't we also consider those customers reporting higher values?
- If so, find a threshold value V and offer to all clients with $v_i \geq V$. Assuming $v_1 \geq v_2 \geq \dots \geq v_n$, optimal threshold found in $O(n^4)$
- Threshold typical in *revenue management*... but not always optimal!
- Sorting by probability or expected value is also sub-optimal
- In fact: **every subset can be optimal — complexity is open**

Threshold strategies are $\frac{1}{2}$ -optimal: LP relaxation

$$V(y) = \sum_{i \in C} v_i \pi_i$$

$$\pi_i = \mathbb{P}[i \text{ accepts and wins}] = y_i \mathbb{E}\left[\frac{1}{1+S_i}\right]$$

$$S_i = \sum_{j \neq i} Y_j \text{ (number of competitors)}$$

$$\implies 0 \leq \pi_i \leq p_i \quad \text{and} \quad \sum_{i \in C} \pi_i \leq 1$$

Threshold strategies are $\frac{1}{2}$ -optimal: LP relaxation

$$V(y) = \sum_{i \in C} v_i \pi_i$$

$$\pi_i = \mathbb{P}[i \text{ accepts and wins}] = y_i \mathbb{E}\left[\frac{1}{1+S_i}\right]$$

$$S_i = \sum_{j \neq i} Y_j \text{ (number of competitors)}$$

$$\implies 0 \leq \pi_i \leq p_i \quad \text{and} \quad \sum_{i \in C} \pi_i \leq 1$$

Consider the relaxation (upper bound)

$$V^* \leq V^{LP} = \max_{0 \leq y_i \leq p_i} \left\{ \sum_{i \in C} v_i y_i : \sum_{i \in C} y_i \leq 1 \right\}$$

and use it to get a $\frac{1}{2}$ -optimal threshold strategy

Linear-time approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^k p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^k p_i & \text{if } i = k + 1 \\ 0 & \text{otherwise} \end{cases}$$

Linear-time approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^k p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^k p_i & \text{if } i = k + 1 \\ 0 & \text{otherwise} \end{cases}$$

- y^{LP} is a randomized strategy equivalent to

$$\begin{cases} \text{select } \{1, \dots, k\} \text{ with probability } [\sum_{i=1}^{k+1} p_i - 1] / p_{k+1} \\ \text{select } \{1, \dots, k + 1\} \text{ with probability } [1 - \sum_{i=1}^k p_i] / p_{k+1} \end{cases}$$

Linear-time approximation algorithm ALG^{LP}

- LP solution in $O(n)$: find largest k with $\sum_{i=1}^k p_i \leq 1$ and set

$$y_i^{LP} = \begin{cases} p_i & \text{if } i \leq k \\ 1 - \sum_{i=1}^k p_i & \text{if } i = k + 1 \\ 0 & \text{otherwise} \end{cases}$$

- y^{LP} is a randomized strategy equivalent to

$$\begin{cases} \text{select } \{1, \dots, k\} \text{ with probability } [\sum_{i=1}^{k+1} p_i - 1] / p_{k+1} \\ \text{select } \{1, \dots, k + 1\} \text{ with probability } [1 - \sum_{i=1}^k p_i] / p_{k+1} \end{cases}$$

- De-randomize in $O(n^2)$: max between $V_{\{1, \dots, k\}}$ and $V_{\{1, \dots, k+1\}}$

Random instances: 10 customers, 200 repetitions

| Algorithm | % opt | Min ratio | Avg ratio | Time |
|----------------|-------|-----------|-----------|--------|
| <i>optimal</i> | 100.0 | 1.0000 | 1.0000 | 1611.2 |
| THRESHOLD | 93.0 | 0.9916 | 0.9998 | 23.5 |
| LP-RELAX | 47.5 | 0.8168 | 0.9771 | 0.4 |

Theorem: ALG^{LP} is a $\frac{1}{2}$ -approximation algorithm

Theorem: ALG^{LP} is a $\frac{1}{2}$ -approximation algorithm

Proof: By Jensen's inequality

$$\mathbb{E}\left[\frac{1}{1+S_i}\right] \geq \frac{1}{1+\mathbb{E}(S_i)} = \frac{1}{1+\sum_{j \neq i} z_j^{LP}} \geq \frac{1}{1+1} = \frac{1}{2}$$

hence

$$V(z^{LP}) = \sum_{i \in C} v_i z_i^{LP} \mathbb{E}\left[\frac{1}{1+S_i}\right] \geq \frac{1}{2} \sum_{i \in C} v_i z_i^{LP}$$

so that

$$V^* \geq V(z^{LP}) \geq \frac{1}{2} V^{LP} \geq \frac{1}{2} V^*.$$

□

Alternative: Hyperbolic relaxation

$$\pi_i \geq \frac{y_i}{1 + \sum_{j \neq i} y_j} \geq \frac{y_i}{1 + \sum_{j \in C} y_j}$$

$$V^* \geq \max_{0 \leq y_i \leq p_i} \frac{\sum_{i \in C} v_i y_i}{1 + \sum_{i \in C} y_i}$$

- *Common-lines problem* in transit equilibrium (Chriqui&Robillard'75)
- Optimum is a threshold strategy
- Linear-time algorithm: $\max_k [v_1 p_1 + \dots + v_k p_k] / [1 + p_1 + \dots + p_k]$
- Also a $\frac{1}{2}$ -approximation algorithm

Improved $\frac{2}{3}$ -approximation

Let $x = \mathbb{P}[\sum_{i \in C} Y_i = 0]$ so that $x + \sum_{i \in C} \pi_i = 1$

Moreover $\pi_i = y_i \mathbb{E}[\frac{1}{1+S_i}]$ with

$$\begin{aligned}\mathbb{E}[\frac{1}{1+S_i}] &\leq 1 \cdot \mathbb{P}[S_i=0] + \frac{1}{2} \cdot \mathbb{P}[S_i > 0] \\ &= \frac{1}{2}(1 + \mathbb{P}[S_i=0]) \\ &= \frac{1}{2}(1 + \frac{x}{1-y_i})\end{aligned}$$

and then $y_i \leq p_i$ implies $\pi_i \leq \frac{p_i}{2}(1 + \frac{x}{1-p_i})$

Hence we get the alternative LP relaxation

$$\begin{aligned} V^* \leq V^{LP_2} &= \max \sum_{i \in C} v_i z_i \\ z_i &\leq \frac{p_i}{2} \left(1 + \frac{x}{1-p_i} \right) \\ x + \sum_{i \in C} z_i &= 1 \\ x, z_i &\geq 0 \end{aligned}$$

Algorithm ALG^{LP_2}

- Find a *basic* optimal solution (z^*, x^*) for LP_2
- Set $y_i^{LP_2} = \frac{2z_i^*}{1 + \frac{x^*}{1-p_i}}$...either 0 or p_i except for one value!
- De-randomize y^{LP_2} to get a set of the form $\{1, \dots, k\}$

Theorem: ALG^{LP_2} is a $\frac{2}{3}$ -approximation

Theorem: ALG^{LP_2} is a $\frac{2}{3}$ -approximation

Proof:
$$V(y) = \sum_{i \in C} v_i y_i \mathbb{E}\left[\frac{1}{1+S_i}\right] \geq \sum_{i \in C} \frac{v_i y_i}{1 + \sum_{j \neq i} y_j}$$

Replacing $y_i^{LP_2}$ we get $V(y^{LP_2}) \geq \sum_{i \in C} v_i z_i^* \gamma_i$ with

$$\gamma_i = \frac{2(1+x^*)}{(3-x^*-2z_i^*)(1+\frac{x^*}{1-p_i})}.$$

Since $V^* \leq \sum_{i \in C} v_i z_i^*$ we need $\gamma_i \geq \frac{2}{3}$. This is obvious if $x^* = 0$. Else, since (z^*, x^*) is a basic solution exactly one of the two inequalities involving z_i^* is tight: if $z_i^* > 0$ then $z_i^* = \frac{p_i}{2}(1 + \frac{x^*}{1-p_i})$ so that

$$\gamma_i = \frac{2(1+x^*)}{(3-p_i-\frac{x^*}{1-p_i})(1+\frac{x^*}{1-p_i})} \geq \frac{2}{3}.$$

□

Random instances: 10 customers, 200 repetitions

| Algorithm | % opt | Min ratio | Avg ratio | Time |
|------------------------|-------|-----------|-----------|--------|
| <i>optimal</i> | 100.0 | 1.0000 | 1.0000 | 1611.2 |
| THRESHOLD | 93.0 | 0.9916 | 0.9998 | 23.5 |
| LP-RELAX | 47.5 | 0.8168 | 0.9771 | 0.4 |
| LP ² -RELAX | 74.0 | 0.9775 | 0.9988 | 21.5 |
| IN-OUT | 99.0 | 0.9918 | 0.9999 | 15.4 |
| SWAP | 100.0 | 1.0000 | 1.0000 | 36.5 |

IN-OUT and SWAP are a local search heuristics that add/remove/swap one element from current strategy to get a maximal increase in revenue.

...complexity remains open

Multiple items

When there are m available items the value of a strategy y is

$$V(y) = \sum_{R \subseteq C} \overbrace{\min \left\{ 1, \frac{m}{|R|} \right\} v(R)}^{\text{REVENUE IF } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in R} y_i}^{\text{PROB THAT } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in C \setminus R} (1 - y_i)}^{\text{PROB THAT } C \setminus R \text{ REJECTS}}$$

Multiple items

When there are m available items the value of a strategy y is

$$\begin{aligned}
 V(y) &= \sum_{R \subseteq C} \overbrace{\min \left\{ 1, \frac{m}{|R|} \right\} v(R)}^{\text{REVENUE IF } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in R} y_i}^{\text{PROB THAT } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in C \setminus R} (1 - y_i)}^{\text{PROB THAT } C \setminus R \text{ REJECTS}} \\
 &= \mathbb{E} \left[\min \left\{ 1, \frac{m}{\sum_{i \in C} Y_i} \right\} \sum_{i \in C} v_i Y_i \right]
 \end{aligned}$$

Multiple items

When there are m available items the value of a strategy y is

$$\begin{aligned}
 V(y) &= \sum_{R \subseteq C} \overbrace{\min \left\{ 1, \frac{m}{|R|} \right\} v(R)}^{\text{REVENUE IF } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in R} y_i}^{\text{PROB THAT } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in C \setminus R} (1 - y_i)}^{\text{PROB THAT } C \setminus R \text{ REJECTS}} \\
 &= \mathbb{E} \left[\min \left\{ 1, \frac{m}{\sum_{i \in C} Y_i} \right\} \sum_{i \in C} v_i Y_i \right] \\
 &= \sum_{i \in C} v_i y_i \mathbb{E} \left[\min \left\{ 1, \frac{m}{1 + S_i} \right\} \right]
 \end{aligned}$$

Multiple items

When there are m available items the value of a strategy y is

$$\begin{aligned}
 V(y) &= \sum_{R \subseteq C} \overbrace{\min \left\{ 1, \frac{m}{|R|} \right\} v(R)}^{\text{REVENUE IF } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in R} y_i}^{\text{PROB THAT } R \text{ ACCEPTS}} \cdot \overbrace{\prod_{i \in C \setminus R} (1 - y_i)}^{\text{PROB THAT } C \setminus R \text{ REJECTS}} \\
 &= \mathbb{E} \left[\min \left\{ 1, \frac{m}{\sum_{i \in C} Y_i} \right\} \sum_{i \in C} v_i Y_i \right] \\
 &= \sum_{i \in C} v_i y_i \mathbb{E} \left[\min \left\{ 1, \frac{m}{1 + S_i} \right\} \right]
 \end{aligned}$$

Problem: find $V^* = \max_{0 \leq y_i \leq p_i} V(y)$

The trivial extension of LP gives an upper bound

$$V^{LP} = \max \left\{ \sum_{i \in C} v_i y_i : \sum_{i \in C} y_i \leq m \text{ and } 0 \leq y_i \leq p_i \right\}$$

which guarantees a constant factor of the optimum:

The trivial extension of LP gives an upper bound

$$V^{LP} = \max \left\{ \sum_{i \in C} v_i y_i : \sum_{i \in C} y_i \leq m \text{ and } 0 \leq y_i \leq p_i \right\}$$

which guarantees a constant factor of the optimum:

$$\begin{aligned} V^* \geq V(y) &= \sum_{i \in C} v_i y_i \mathbb{E}[\min\{1, \frac{m}{1+S_i}\}] \\ &\geq \rho(y) \sum_{i \in C} v_i y_i = \rho(y) V^{LP} \geq \rho(y) V^* \end{aligned}$$

$$\rho(y) = \mathbb{E}[\min\{1, \frac{m}{1+S}\}] \geq 1 - \frac{1}{m+1} \sqrt{1 + \text{Var}(S)} \mathbb{P}(S \geq m)$$

find lower bound for $\rho(y)$...

Asymptotically optimal bound

Since $\text{Var}(S) = \sum_i y_i(1 - y_i) \leq m$ and $\mathbb{P}(S \geq m) \leq 1$, we get

$$\rho(y) \geq 1 - \frac{1}{\sqrt{m+1}}$$

hence

Theorem: $V(y^{LP})$ is within a factor $\frac{1}{\sqrt{m+1}}$ from optimal

...and not better!

Sharper bound

$$\rho(y) \geq 1 - \frac{\sqrt{1+\sigma^2}}{m+1} \min\left\{1, \frac{1}{2} + \frac{M}{\sigma}\right\}$$

$$M = \max_{u \geq 0} \sqrt{2u} e^{-2u} \sum_{k=0}^{\infty} \left(\frac{u^k}{k!}\right)^2 \sim 0.46882235549939533$$

“if the mean number of successes in n independent heterogeneous Bernoulli trials is an integer m then the median is also m ”

(Jogdeo-Samuels'62, Siegel'01)

$$\mathbb{P}[S \geq m] \leq \frac{1}{2} + \mathbb{P}[S = m] \leq \frac{1}{2} + \frac{M}{\sigma}$$

(latter estimate from C+Vaisman'08)

Extensions — Buy by Bulk

Dean, Goemans, Vondrak 04

- The difference is that customer $i \in C$ demands $s_i \geq 1$ units
- Problem is now NP-hard: for $p_i = 1$ we get knapsack
- Computing $V(y)$ is actually #P-complete
(e.g. count the number of perfect matchings in a bipartite graph)

Extensions — Multistage

- Medium-term: address offers in k successive rounds.
- Find an adaptive online algorithm that computes an optimal subset of clients to address the offer in the next round.

Extensions — Pricing

- Personalized contact (email): price can be different for each client
- Pbb of accepting is a decreasing function of price $p_i : \mathbb{R}_+ \rightarrow [0, 1]$
- Select a price for each customer to maximize expected revenue