

Optimal selection of customers for a last-minute offer

Roberto Cominetti* José R. Correa† Thomas Rothvoß‡ Jaime San Martín§

Abstract

We analyze a short-term revenue optimization problem involving the targeting of customers for a promotion in which a finite number of perishable items are sold on a last-minute offer. The goal is to select the subset of customers to whom the offer will be made available in order to maximize the expected return. Each client replies with a certain probability and reports a specific value that may depend on the customer type, so that the selected subset has to balance the risk of not selling all items with the risk of assigning an item to a low value customer.

We show that *threshold strategies*, which select all those clients with values above a certain optimal threshold, may fail to achieve the maximal revenue. However, using a linear programming relaxation, we prove that they attain a constant factor of the optimal value. Specifically, the achieved factor is $1/2$ when a single item is to be sold, and approaches 1 as the number of available items grows to infinity. Also, for the single item case, we propose an upper bound based on a sharper linear relaxation that allows us to obtain a threshold strategy achieving at least $2/3$ of the optimal revenue. Moreover, although the complexity status of the problem is open, we provide a polynomial time approximation scheme for the single item case.

Keywords: promotional sale, linear programming relaxations, approximation algorithms.

1 Introduction

In this paper we consider a short-term revenue optimization problem that involves selecting a set of customers for a promotion in which a finite number of perishable items are sold on a last-minute offer. The goal is to maximize the total expected return by choosing the subset of customers to whom the offer will be made available, taking into account that each client replies with a certain probability and reports a specific value that may depend on the customer type. Any item not sold is either lost or derived to a secondary market at a fixed price which sets a reference value from which we measure a customer's surplus value.

As a motivation example consider the situation faced by an airline offering *discount business upgrades* on a flight close to its departure date, say two days ahead. The offer may be sent to some or all of the passengers that hold a current booking in economy class for that flight. The airline values its clients differently based on personal characteristics such as past travel behavior, propensity to buy an upgrade at regular price, level of fidelization, or other specific demographic data. However, because of lack of time it is not possible to make the offer on a one-by-one basis from higher to lower valued customers and they must all be contacted at once by e-mail, after which the items will be assigned according to a first-come-first-served discipline.

*Departamento de Ingeniería Matemática, Universidad de Chile, Santiago, Chile (rcominet@dim.uchile.cl). Partially supported by Instituto Milenio Sistemas Complejos de Ingeniería.

†Departamento de Ingeniería Industrial, Universidad de Chile, Santiago, Chile (jcorrea@dii.uchile.cl). Partially supported by Instituto Milenio Sistemas Complejos de Ingeniería.

‡Institute of Mathematics EPFL, Lausanne, Switzerland (thomas.rothvoss@epfl.ch).

§Departamento de Ingeniería Matemática and Centro de Modelamiento Matemático, Universidad de Chile, Santiago, Chile (jsanmart@dim.uchile.cl). Supported by FONDAPE Matemáticas Aplicadas - CONICYT.

A similar situation happens with the business class upgrades sold *at the gate* by several airlines. Available business class seats are offered at a certain price to a subset of the customers which are waiting for boarding, who may be discriminated according to the fare class they bought. Although the price is flat, the airline values its clients differently as consistently making upgrades available for cheaper fares disincentives the sales for more expensive ones. If however the upgrades are only made available to the highest fare class, the airline risks of not selling enough upgrades. One can imagine similar situations in the car or computer industries where a company wants to put on sale some old units currently held in stock, before a new model is released to the market.

Intuitively, the promotion should include enough customers to have a high probability of selling all the items. However, if too many customers are included one may face an excess of demand and end up selling an item to a low revenue client who displaces a higher valued customer responding later. Handling this type of tradeoffs is a basic issue in revenue management problems. A natural policy is to select all customers whose values are above a certain threshold which is chosen so as to maximize the expected revenue. Such strategies are indeed optimal if low revenue clients react to the promotion before higher revenue clients, but may fail to be optimal otherwise. In our short term setting we assume that there is no correlation between the order in which customers react and their value. In this case we show that a threshold policy, although not optimal, attains a guaranteed fraction of the optimal revenue. Moreover this fraction becomes asymptotically close to one as the number of items grows.

In practice, the estimation of the acceptance probabilities and the values of clients may be based on historic data accumulated from previous promotional sales, possibly by grouping the clients into classes or categories according to different demographic and behavioral characteristics. Since such estimates may be imprecise it is worth noting that, although small errors may significantly change the optimal solution, the optimal objective value varies continuously with the input data. Furthermore, threshold policies are robust under calibration errors, as they only depend on the order of clients rather than on specific values. Then our approximation results for these type of policies hold even under noisy data.

Our setting differs significantly from classic revenue management approaches (we refer the reader to [10] for a detailed treatment). Probably the most relevant difference comes from our short notice assumption. For instance, without this assumption the so-called nested policies, i.e., threshold policies applied over several rounds, turn out to be optimal [1]. On the other hand, although sampling targets from a population for a direct mail campaign is an issue already considered in marketing [2], our model deviates from the marketing literature since the short notice assumption does not allow to take a dynamic approach. In this sense note that the first-come-first-served discipline appears to be reasonable in a short notice setting, as it is indeed the right model whenever the response times of the customers are independent and identically distributed. This contrasts with rationing policies that yield better revenues when more time is available to sell the items.

The paper is structured as follows. Section §2 gives a mathematical description of the problem, which leads to a basic, though unexplored, discrete optimization problem. We present in §2.2 counterexamples to some natural properties that one might expect such as the non-optimality of threshold policies, while in §3 we show that these type of solutions lead to algorithms with explicit bounds on the approximation guarantees. In particular we describe a trivial heuristic with a $\frac{1}{2}$ -approximation guarantee for the single item case and which becomes asymptotically optimal when the number of items on sale goes to infinity. In §4 we discuss a sharper LP relaxation and derive an improved $\frac{2}{3}$ -approximation algorithm for the single item case, which is also a threshold strategy. Then, in §5, we deviate from threshold strategies and provide an algorithm which takes any $\varepsilon > 0$ and outputs a $(1 + \varepsilon)$ -approximate solution in time which is polynomial for fixed ε , that is to say a polynomial time approximation scheme (PTAS). We conclude in §6 pointing out some directions

for further research, while in the Appendix §7 we present statistics on the numerical performance of the proposed algorithms and their comparison with two additional heuristics.

2 Problem formulation

A finite number of m identical perishable items is to be offered on a promotional sale among a finite set of customers $C = \{1, \dots, n\}$. Each customer $i \in C$ is characterized by a probability $p_i \in [0, 1]$ of accepting the offer, and a value $v_i \geq 0$ which he reports when buying an item. This value represents the net expected revenue, namely, the promotional price minus the expected value of the customer if he is excluded from the promotion, accounting for the possibility that he would buy an item anyway at the regular price. We assume that any unsold item is lost, as for instance in the case of empty seats on a given flight close to its departure date. We also suppose $\sum p_i \geq m$ and that customers have been pre-ordered by decreasing values so that $v_1 \geq v_2 \geq \dots \geq v_n$. The problem is to choose the set of customers $A \subseteq C$ to whom the promotional sale will be made available. If m is large more clients should be included while for small m one should be more selective to avoid high-revenue customers being displaced by low-revenue clients. The goal is to balance the risk of displacement and the risk of not selling all the items, maximizing the total expected return.

Note that we neither assume that the values v_i are similar in any way, nor that they are correlated with the probabilities p_i . Thus the model is also suitable to incorporate other customer-specific characteristics in other type of promotional sales, including regional considerations, loyalty issues, dispatching costs, among others. In some situations, it may be reasonable to assume that $p_i > p_j$ whenever $v_i > v_j$. If this is the case, it is easy to observe that threshold policies (i.e., policies selecting all customers reporting a value above a certain threshold) are optimal. However, this assumption is not necessarily the most natural, since v_i is the value of the client from the point of view of the company, not the valuation from the point of view of the customer. For instance, consider again the business upgrade example given in the introduction, and assume the value of a client for the company is given by the discounted price of the upgrade minus the regular price times the probability a customer buys at regular price. In such a situation, a customer with low propensity to buy (small p_i) will also be less likely to induce dilution (i.e., his propensity to buy at regular price if he does not receive the offer will also be small), and therefore the value of such a client from the point of view of the company will be higher.

2.1 The optimization problem

To state the problem precisely we introduce decision variables $x_i \in [0, 1]$ which represent the probability with which we make the offer available to customer i . This approach involves a randomized choice of the customers and covers the possibility of an include/exclude policy represented by the extreme values $x_i = 1$ or $x_i = 0$. Alternatively we consider as decision variable the value $y_i = p_i x_i$ which represents the probability for customer i to be a *respondent*, that is to say, a client who receives the offer and is willing to accept it. Consider the Bernoulli random variables

$$Y_i = \begin{cases} 1 & \text{if customer } i \text{ is a respondent} \\ 0 & \text{otherwise} \end{cases}$$

with $\mathbb{P}(Y_i = 1) = y_i$, and let $R = \{i \in C : Y_i = 1\}$ be the random set of respondents. If $|R| \leq m$ all the demand can be satisfied yielding a total revenue of

$$v_R = \sum_{j \in R} v_j.$$

Otherwise, since only m items are available we must choose m customers from R to assign the items. If this is done according to a *first-come-first-served* discipline and we assume no correlation between the (p_i, v_i) 's and the reaction time, every subset $B \subseteq R$ of m clients is equally likely to occur with probability $q = 1/\binom{|R|}{m}$, and then the expected value for R is

$$v_R = \sum_{B \subseteq R, |B|=m} (\sum_{j \in B} v_j) q = \frac{m}{|R|} \sum_{j \in R} v_j.$$

Both cases may be recombined in the formula $v_R = \min\{1, \frac{m}{|R|}\} \sum_{j \in R} v_j$, with the convention $v_R = 0$ when $R = \phi$. Finally, assuming that customers act independently, the total expected revenue is

$$V(y) = \sum_{R \subseteq C} v_R p_R(y) = \mathbb{E} \left[\min\{1, \frac{m}{\sum_{j \in C} Y_j}\} \sum_{j \in C} v_j Y_j \right] \quad (1)$$

where $p_R(y) = \prod_{j \in R} y_j \prod_{j \notin R} (1 - y_j)$ is the probability of observing the respondent set R . The optimization problem to be solved is therefore

$$(P) \quad V^* = \max_{0 \leq y_i \leq p_i} V(y).$$

Although the variables y_i are continuous, problem (P) has a discrete nature: since $V(y)$ is affine with respect to each y_i it follows that (P) has an extremal optimal solution with $y_i \in \{0, p_i\}$. Indeed, if y^* is optimal, the slope corresponding to any “fractional” variable $0 < y_i^* < p_i$ must be zero so that rounding up to p_i or down to 0 indifferently does not change the value and yields an extremal solution and a corresponding optimal subset of customers $A^* = \{i \in C : y_i^* = p_i\}$ with no randomization involved. More generally, to any feasible y we may associate an extremal solution $[y]$ with $V([y]) \geq V(y)$ by using the following *monotone rounding* procedure: consider each fractional variable $0 < y_i < p_i$ sequentially and compare the alternative solutions with $y_i = p_i$ and $y_i = 0$, updating y to the best of both. We will denote $A[y]$ the support of this solution, that is, the set of those $i \in C$ with $[y]_i > 0$.

It follows that, when C is a small set, (P) may be solved by *brute force* considering all possible subsets $A \subseteq C$ and comparing the values $V_A = V(y^A)$ with $y_i^A = p_i$ for $i \in A$ and $y_i^A = 0$ otherwise. This is feasible provided that $V(y)$ can be computed efficiently. The sum in (1) has exponentially many terms and is not well suited for direct computation. An alternative expression may be derived considering the random variables $S = \sum_{j \in C} Y_j$ and $S_i = \sum_{j \neq i} Y_j$ which allows us to write

$$V(y) = \sum_{i \in C} v_i \pi_i(y), \quad (2)$$

$$\pi_i(y) = \mathbb{E} \left[\min\{1, \frac{m}{S}\} Y_i \right] = y_i \mathbb{E} \left[\min\{1, \frac{m}{1+S_i}\} \right], \quad (3)$$

where $\pi_i(y)$ represents the probability that customer $i \in C$ gets an item (*i.e.* he receives an offer, he accepts it, and he is lucky enough to get one of the available items). The distribution of each S_i may be computed by convolution in $O(n^2)$ arithmetic operations (see [5]), so that (2)-(3) provide an effective way to compute $V(y)$ in $O(n^3)$.

2.2 Conjectures and counterexamples

In order to reduce the effort involved in solving (P) by brute force, one may seek for structural properties to limit the search to a smaller range of subsets. Unfortunately, most intuitive conjectures turn out to be false. For instance, since high values v_i are preferable it is natural to expect an optimal set of the form $A^* = \{1, \dots, k\} = \{i : v_i \geq \bar{v}\}$ for a certain threshold value \bar{v} , a common

policy in revenue management [8, 10]. In our setting this turns out to be sub-optimal as shown by an elementary example for $m = 1$ items and $n = 3$ customers with characteristics $p_1 = p_2 = 0.5$, $p_3 = 1$, $v_1 = 2$, $v_2 = 1$, $v_3 = 0.9$, for which the optimal solution is $A^* = \{1, 3\}$. Roughly speaking, a customer is valuable only if his contribution to reducing the risk of not selling all the items compensates the losses produced when he displaces a higher valued client. In the table above the second player is not valuable enough but the third is, despite the fact that $v_2 > v_3$. This occurs since p_2 is too small while p_3 is large enough, showing that the decision to include a customer depends not only on the values but also the probabilities of all clients.

In general the highest value customer is always in the optimal set as he cannot displace any higher valued demand. Also using (2) it follows that if $i \in A^*$ then A^* contains all customers j satisfying $v_j > v_i$ and $p_j \geq p_i$ simultaneously. Observe that this implies that if p_i 's and v_i 's are positively correlated then the optimal solution is a threshold strategy, and thus can be found efficiently. However, except for such simple properties, the structure of the optimal set has little to do with the order of the v_i 's. As a matter of fact, any set containing the highest value customer can be optimal. More precisely let $\bar{A} \subseteq \mathbb{Z}_+$ be any finite set with $1 \in \bar{A}$. Then one can find an instance with $m = 1$, in which customers are strictly ordered by decreasing values $v_1 > v_2 > \dots > v_n$, and where the optimal solution is exactly $A^* = \bar{A}$. In order to build such an instance let $C = \{1, \dots, n\}$ with n to be set later on, and take momentarily $v_1 = 10, p_1 = 0.1$ and $v_i = 1, p_i = \varepsilon$ for $i = 2, \dots, n$. Now, any subset A including customer 1 together with k additional clients has expected revenue

$$V_\varepsilon(k) = 1 - \frac{0.9}{(k+1)\varepsilon} [(1 - \varepsilon)^{k+1} - 1 + (k + 1)(1 - \varepsilon)^k \varepsilon].$$

Considering the monotone change of variables $y = (1 - \varepsilon)^{k+1}$ we have $V_\varepsilon(k) = 1 - \frac{0.9 \ln(1-\varepsilon)}{\varepsilon} g(y)$, where $g(y) = (y - 1) / \ln y + y \frac{\varepsilon}{(1-\varepsilon) \ln(1-\varepsilon)}$ is strictly concave and attains its maximum at a unique point $y_\varepsilon \in (0, 1)$. It turns out that $V_\varepsilon(k)$ as a function of the continuous variable k has a unique maximum, which is characterized by the first order optimality condition. More explicitly, taking λ such that $e^{-\lambda} = 1 - \varepsilon$ and setting $x = k + 1$, this optimality condition reads $e^{\lambda x} = \lambda x (e^{\lambda x} - 1) + 1$. We observe that for any $x > 1$, by taking $\lambda = 1/x$ the right hand side of this equation is larger, while for λ big the left hand side dominates. Hence, there exists λ for which the equality holds at x . In particular, for $x = |\bar{A}|$ one may find $\varepsilon > 0$ such that the maximum of $V_\varepsilon(k)$ is attained precisely at $\bar{k} = |\bar{A}| - 1$. Now, using this fact and taking $n > \max\{i : i \in \bar{A}\}$ it follows that any subset of C which contains customer 1 and \bar{k} additional customers is optimal. Moreover, if we decrease the probabilities $p_i = \varepsilon$ to $p_i < \varepsilon$ for all $i \in C \setminus \bar{A}$, we get that the only optimal solution is \bar{A} . Finally we may perturb the values v_i so that $v_1 > v_2 > \dots > v_n$, with a perturbation small enough to ensure that \bar{A} remains the unique optimal solution.

Given the previous counter-example, one might think that sorting the customers according to an alternative criteria such as expected revenue $p_i v_i$ could be more appropriate (except for the highest valued customer who is always included in A^*). However, not only the optimal set can differ significantly from any set complying with the order of $p_i v_i$'s, but also any such solution may be arbitrarily poor. Indeed, let $\varepsilon > 0$ and consider the instance with $C = C_1 \cup C_2 = \{1, \dots, n\}$ where n is sufficiently large. Take $v_i = 1, p_i = 1$ for $i \in C_1$ and $v_i = 1/\varepsilon, p_i = \varepsilon$ for $i \in C_2$, and slightly perturb the values v_i so that the customers in C_1 have slightly higher $p_i v_i$. Clearly, for large $|C_2|$, the optimal solution has value V^* arbitrarily close to $1/\varepsilon$. On the other hand, if in addition $|C_1| \gg |C_2|$, the best solution complying with the order of $p_i v_i$'s has value arbitrarily close to 1.

3 A simple approximation algorithm

Although the computational complexity of problem (P) is open, the examples in §2.2 suggest that an optimal solution may have a complicated structure. These observations motivate the following question: Can the best $\{1, \dots, k\}$ strategy be far away from V^* ? In the sequel we answer this question in the negative by providing constant factor approximation algorithms returning solutions of this form. These algorithms run in polynomial-time and have an approximation guarantee that improves and becomes asymptotically optimal as the number of items m increases. Since $V(y)$ can be computed in $O(n^3)$, the best threshold solution for the form $A^* = \{1, \dots, k\}$ can be found in $O(n^4)$ arithmetic operations by running the following:

ALGORITHM 1 — MAX- k
compute $V_k = V(p_1, \dots, p_k, 0, \dots, 0)$ for $k = 1, \dots, n$
choose the best V_k and set $A_1^* = \{1, \dots, k\}$.

We will measure how far the solution A_1^* may be from the optimal value by comparing to an even simpler method which requires only 2 function evaluations with an $O(n^3)$ complexity. Surprisingly, this trivial heuristic computes a threshold type solution that becomes asymptotically optimal when the number of items m is large:

ALGORITHM 2 — ADD- m
find the largest k with $p_1 + \dots + p_k \leq m$
let $\bar{z} = (p_1, \dots, p_k, \rho, 0, \dots, 0)$ with $\rho = m - \sum_{i=1}^k p_i$
set $A_2^* = A[\bar{z}]$ the support of the monotone rounding of \bar{z} .

Note that a randomized variant of this method which computes \bar{z} and then selects $\{1, \dots, k\}$ with probability $1 - \rho/p_{k+1}$ and $\{1, \dots, k+1\}$ with probability ρ/p_{k+1} , gives a solution whose expected value satisfies all the guarantees we prove for Algorithm 2. Since no monotone rounding is involved, this variant runs in $O(n)$ without ever computing the objective function!

In order to estimate the quality of the solutions A_1^* and A_2^* , we consider the following relaxation of (P) . We observe from (3) that $\pi_i(y) \leq y_i$ and $\sum_i \pi_i(y) = \mathbb{E}[\min\{S, m\}] \leq m$, so that V^* may be bounded from above by the linear program

$$V^* \leq \max_z \left\{ \sum_i v_i z_i : z_i \in [0, p_i]; \sum_i z_i \leq m \right\} = \sum_i v_i \bar{z}_i = \bar{V}$$

where the optimum \bar{z} is obtained by putting as much weight as possible on higher valued customers, which is precisely the solution computed in ADD- m . Clearly $V(\bar{z}) \leq V^* \leq \bar{V}$ and since the set A_2^* corresponds to the monotone rounding $[\bar{z}]$ we deduce

$$V(\bar{z}) \leq V_{A_2^*} \leq V_{A_1^*} \leq V^* \leq \bar{V}. \quad (4)$$

A lower bound showing that $V(\bar{z})$ is not too far from the optimal value V^* is obtained by letting

$$\rho(y) = \mathbb{E} \left[\min \left\{ 1, \frac{m}{1+S} \right\} \right],$$

so that $\pi_i(y) \geq y_i \rho(y)$. Hence $V(y) \geq \rho(y) \sum_i v_i y_i$ from which we get

Lemma 1 $V(\bar{z}) \geq \rho(\bar{z}) \bar{V}$.

3.1 The single item case

A direct consequence of (4) and Lemma 1 is that \bar{z} yields a $\frac{1}{2}$ -approximate solution if $m = 1$.

Proposition 2 *If* $m = 1$ *then* $V(\bar{z}) \geq \frac{1}{2} V^*$ *and* A_1^*, A_2^* *are* $\frac{1}{2}$ -*approximate solutions for* (P) .

Proof. It suffices to note that for \bar{z} we have $\mathbb{E}(S) = 1$ and then $\rho(\bar{z}) = \mathbb{E}(\frac{1}{1+S}) \geq \frac{1}{1+\mathbb{E}(S)} = \frac{1}{2}$. Hence $V(\bar{z}) \geq \frac{1}{2}\bar{V}$ which combined with (4) gives the desired conclusion. \square

The bound of $\frac{1}{2}$ is tight for $m = 1$. For instance, if $v_1 = 1$, $p_1 = \varepsilon$ and $v_2 = 0$, $p_2 = 1 - \varepsilon$, the optimal solution is $y^* = (p_1, 0)$ with value $V^* = \varepsilon$, while the relaxed problem gives $\bar{z} = (p_1, p_2)$ with $V(\bar{z}) = \frac{1}{2}[\varepsilon + \varepsilon^2]$. The ratio $V(\bar{z})/V^*$ clearly tends to $\frac{1}{2}$ as $\varepsilon \downarrow 0$. Thus our analysis is tight.

3.2 The multiple item case

In order to prove the asymptotic optimality of \bar{z} when $m \rightarrow \infty$, we establish a lower bound for $\rho(\bar{z})$ that converges to 1. To this end we note that using the Cauchy-Schwartz inequality we get

$$\begin{aligned} \rho(\bar{z}) &= 1 - \mathbb{E}\left[\frac{1+S-m}{1+S} \mathbb{1}_{\{S \geq m\}}\right] \\ &\geq 1 - \frac{1}{m+1} \mathbb{E}[(1+S-m) \mathbb{1}_{\{S \geq m\}}] \\ &\geq 1 - \frac{1}{m+1} \sqrt{\mathbb{E}[(1+S-m)^2]} \sqrt{\mathbb{P}(S \geq m)}. \end{aligned}$$

Since $\mathbb{E}(S - m) = 0$ we have $\mathbb{E}[(1+S-m)^2] = 1 + \bar{\sigma}^2$ where $\bar{\sigma}^2 = \text{Var}(S) = \sum_i \bar{z}_i(1 - \bar{z}_i)$, and then

$$\rho(\bar{z}) \geq 1 - \frac{\sqrt{1+\bar{\sigma}^2}}{m+1} \sqrt{\mathbb{P}(S \geq m)}. \quad (5)$$

Now, $\bar{\sigma}^2 \leq m$ and $\mathbb{P}(S \geq m) \leq 1$ imply the lower bound $\rho(\bar{z}) \geq 1 - \frac{1}{\sqrt{m+1}}$, showing that $V(\bar{z})$ is asymptotically optimal for large m . A sharper bound can be obtained using [7, Theorem 3.2] which states that “if the mean number of successes in n independent heterogeneous Bernoulli trials is an integer m then the median is also m ” (see [11]).

Proposition 3 *Let $\bar{\sigma}^2 = \sum_i \bar{z}_i(1 - \bar{z}_i)$. Then $\rho(\bar{z}) \geq 1 - \frac{1}{\sqrt{m+1}} \sqrt{\frac{1}{2} + \frac{1}{2\sqrt{m}}}$.*

Proof. Since $\mathbb{E}(S) = m \in \mathbb{N}$, the median of S is also m , that is to say $\mathbb{P}(S > m) \leq \frac{1}{2} \leq \mathbb{P}(S \geq m)$, and therefore $\mathbb{P}(S \geq m) = \mathbb{P}(S > m) + \mathbb{P}(S = m) \leq \frac{1}{2} + \mathbb{P}(S = m)$. Now, according to [4, 12] we have $\mathbb{P}(S = m) \leq \frac{M}{\bar{\sigma}}$ with $M = \max_{u \geq 0} \sqrt{2u} e^{-2u} \sum_{k=0}^{\infty} \left(\frac{u^k}{k!}\right)^2 < \frac{1}{2}$, so that $\mathbb{P}(S \geq m) \leq \min\{1, \frac{1}{2} + \frac{1}{2\bar{\sigma}}\}$, which plugged into (5) yields

$$\rho(\bar{z}) \geq 1 - \frac{\sqrt{1+\bar{\sigma}^2}}{m+1} \min\{1, \sqrt{\frac{1}{2} + \frac{1}{2\bar{\sigma}}}\}.$$

Since the map $\sigma \mapsto (1+\sigma^2) \min\{1, \frac{1}{2} + \frac{1}{2\sigma}\}$ is increasing, the conclusion follows from $\bar{\sigma} \leq \sqrt{m}$. \square

Proposition 3 implies that the solutions A_1^* and A_2^* are asymptotically optimal with rate of convergence of order $1/\sqrt{m}$. This rate cannot be improved because of the central limit theorem. Indeed, consider a large instance of n identical clients with $v_i = 1$ and $p_i = p$. The optimal solution for any number of items m is to include all customers $A^* = C$, with optimal value $V^* \sim m$ for n large enough. Now suppose that $p = \frac{m}{k}$ with k integer so that $A_2^* = \{1, \dots, k\}$ which according to (1) gives a value

$$V_{A_2^*} = \mathbb{E}[\min\{S, m\}] = m + \mathbb{E}[\min\{\frac{S-m}{\sqrt{m(1-p)}}, 0\}] \sqrt{m(1-p)}$$

where S has binomial distribution $B(k, p)$ with $\mathbb{E}(S) = m$ and $\text{Var}(S) = m(1-p)$. If we now consider a case with p fixed and m, k large, the central limit theorem implies that $V_{A_2^*} \sim m[1 - \sqrt{\frac{1-p}{2\pi m}}]$ which underestimates the optimal value by a factor of the order $\frac{1}{\sqrt{m}}$.

3.3 An alternative approach for the single item case

Both solutions A_1^* and A_2^* provide a lower bound for the optimal value V^* . We proved in Proposition 2 that for $m = 1$ the value $V_{A_2^*}$ is at least $\frac{1}{2}\bar{V}$ so that $V^* \in [\frac{1}{2}\bar{V}, \bar{V}]$. An improved lower bound (and a corresponding approximate solution) can be obtained using (3) and noting that $\pi_i(y) = y_i \mathbb{E}[\frac{1}{1+S_i}] \geq \frac{y_i}{1+\mathbb{E}(S_i)} = \frac{y_i}{1+\sum_{j \neq i} y_j} \geq \frac{y_i}{1+\sum_{j \in C} y_j}$. Therefore $V^* \geq \tilde{V}$, with $\tilde{V} = \max_{0 \leq y_i \leq p_i} (\sum_{i \in C} v_i y_i) / (1 + \sum_{i \in C} y_i)$. The latter is closely related to the *common lines problem* in transit network optimization [3] and can be easily solved by a linear time algorithm:

ALGORITHM 3 — MAX-AVG
find k *with* $[v_1 p_1 + \dots + v_k p_k] / [1 + p_1 + \dots + p_k]$ *maximal*
set $A_3^* = \{1, \dots, k\}$.

The set found by this algorithm satisfies $\tilde{V} \leq V_{A_3^*} \leq V^*$. Note that for each z feasible in the relaxation \bar{V} we have

$$\sum_{i \in C} v_i z_i \leq 2 \frac{\sum_{i \in C} v_i z_i}{1 + \sum_{i \in C} z_i} \leq 2\tilde{V},$$

from which we get $\frac{1}{2}\bar{V} \leq \tilde{V}$ so that \tilde{V} is a sharper lower bound and then $V^* \in [\tilde{V}, \bar{V}]$. It follows that A_3^* is a $\frac{1}{2}$ -approximate solution for (P) as well.

4 A sharper relaxation

We now derive an improved LP relaxation of (P) , which follows from a nonlinear nonconvex formulation, and use it to obtain a $\frac{2}{3}$ -approximation algorithm for the single item case. Using the expression $V(y) = \sum_{R \subseteq C} v_R p_R(y)$ given in (1) and introducing new variables $x_R = p_R(y)$ we have

$$\begin{aligned} V^* &= \max \sum_{R \subseteq C} v_R x_R \\ &0 \leq y_i \leq p_i \text{ for all } i \in C \\ &x_R = \prod_{j \in R} y_j \prod_{j \notin R} (1 - y_j) \text{ for all } R \subseteq C. \end{aligned}$$

The x_R 's are equivalently defined by the constraints $x_R = x_{R \setminus \{i\}} \frac{y_i}{1 - y_i}$ for all $i \in R$ together with $\sum_{R \subseteq C} x_R = 1$, so that the problem becomes

$$\begin{aligned} V^* &= \max \sum_{R \subseteq C} v_R x_R \\ &0 \leq y_i \leq p_i \text{ for all } i \in C \\ &x_R = x_{R \setminus \{i\}} \frac{y_i}{1 - y_i} \text{ for all } R \subseteq C, i \in R \\ &\sum_{R \subseteq C} x_R = 1 \end{aligned}$$

which may then be relaxed into the following linear program

$$\begin{aligned} V^* &\leq V_{LP} = \max \sum_{R \subseteq C} v_R x_R \\ &0 \leq x_R \leq x_{R \setminus \{i\}} \frac{p_i}{1 - p_i} \text{ for all } R \subseteq C, i \in R \\ &\sum_{R \subseteq C} x_R = 1. \end{aligned}$$

Though there are instances for which $V^* < V_{LP}$, the following property of extremal solutions for V_{LP} provides intuition on the structure of the above linear program.

Lemma 4 *Let x be a basic solution for V_{LP} . Then for all $R \neq \phi$ such that $x_R > 0$ we have*

$$x_R = x_\phi \prod_{i \in R} \frac{p_i}{1 - p_i}. \quad (6)$$

Proof. The inequalities $x_R \leq x_{R \setminus \{i\}} \frac{p_i}{1-p_i}$ inductively imply $x_R \leq x_\phi \prod_{i \in R} \frac{p_i}{1-p_i}$ and then, since the sum of the x_R 's is 1, it follows that $x_\phi > 0$.

Let us build a graph $G = (N, E)$ with one vertex $n_R \in N$ for each $R \subseteq C$ with $x_R > 0$, and an edge $(n_R, n_{R \setminus \{i\}}) \in E$ iff $i \in R$ and $x_R = x_{R \setminus \{i\}} \frac{p_i}{1-p_i}$. We claim that G is connected, from which (6) follows at once. If this was not the case we could find a nontrivial partition of N into $L \cup U$ with $n_\phi \in L$, such that $x_R < x_{R \setminus \{i\}} \frac{p_i}{1-p_i}$ when $n_R \in U$ and $n_{R \setminus \{i\}} \in L$. Let $\varepsilon, \delta > 0$ be such that $\varepsilon \sum_{R \in U} x_R = \delta \sum_{R \in L} x_R$, and small enough so that \bar{x} and \tilde{x} defined as

$$\bar{x}_R = \begin{cases} x_R(1 - \varepsilon), & \text{for all } R \in U \\ x_R(1 + \delta), & \text{for all } R \in L, \end{cases} \quad \text{and} \quad \tilde{x}_R = \begin{cases} x_R(1 + \varepsilon), & \text{for all } R \in U \\ x_R(1 - \delta), & \text{for all } R \in L, \end{cases}$$

are feasible for LP. It follows immediately that $x = \frac{1}{2}\bar{x} + \frac{1}{2}\tilde{x}$, and thus x is not an extreme point. This contradiction shows that G is connected which completes the proof. \square

Although V_{LP} seems to be a fairly tight upper bound on V^* , it may be hard to compute since it has exponentially many variables and constraints. We may however derive a weaker relaxation for the single item case, which yields a $\frac{2}{3}$ -approximation algorithm. Let us rewrite the LP as

$$\begin{aligned} V_{LP} &= \max \sum_{i \in C} v_i \sum_{R \ni i} \frac{x_R}{|R|} \\ &0 \leq x_R \leq p_i(x_{R \setminus \{i\}} + x_R) \text{ for all } R \subseteq C, i \in R \\ &\sum_{R \subseteq C} x_R = 1. \end{aligned}$$

Letting $z_i = \sum_{R \ni i} \frac{x_R}{|R|}$ we get $\sum_{i \in C} z_i = 1 - x_\phi$, while the inequalities $x_R \leq p_i(x_{R \setminus \{i\}} + x_R)$ yield

$$z_i \leq \sum_{R \ni i} \frac{p_i(x_{R \setminus \{i\}} + x_R)}{|R|} \leq p_i(x_\phi + x_{\{i\}}) + \frac{p_i}{2} \sum_{R \ni i, |R| \geq 2} (x_{R \setminus \{i\}} + x_R).$$

Since $\sum_R x_R = 1$ the last sum is equal to $1 - x_\phi - x_{\{i\}}$, so that $z_i \leq (p_i/2)(x_\phi + x_{\{i\}} + 1) \leq (p_i/2) \left[\frac{x_\phi}{1-p_i} + 1 \right]$. Combining these inequalities and renaming the variable x_ϕ as x we get

$$\begin{aligned} V_{LP} &\leq V_{LP_2} = \max \sum_{i \in C} v_i z_i \\ &z_i \leq \frac{p_i}{2} \left[\frac{x}{1-p_i} + 1 \right] \text{ for all } i \in C \\ &\sum_{i \in C} z_i = 1 - x \\ &z_i, x \geq 0 \end{aligned}$$

from which we derive the following alternative approximation algorithm

ALGORITHM 4 — LP-RELAX

find a basic optimal solution (z^*, x^*) for V_{LP_2} and let $y_i = 2z_i^* / (\frac{x^*}{1-p_i} + 1)$

set $A_4^* = A[y]$ the support of the monotone rounding of y .

Note that the solution given by this algorithm is also of threshold type $A_4^* = \{1, \dots, k\}$. Indeed, once the optimal x^* is known, the optimal values z_i^* are obtained by putting as much weight as possible on high-value customers, saturating the constraints.

Proposition 5 *When $m = 1$ we have $V_{A_4^*} \geq \frac{2}{3}V^*$.*

Proof. Since $V_{A_4^*} \geq V(y)$ and $V_{LP_2} \geq V^*$ it suffices to show $V(y) \geq \frac{2}{3}V_{LP_2}$. To this end we write

$$V(y) = \sum_{i \in C} v_i y_i \mathbb{E} \left(\frac{1}{1 + \sum_{j \neq i} Y_j} \right) = \sum_{i \in C} v_i z_i^* \gamma_i$$

with $\gamma_i = 2\mathbb{E}(\frac{1}{1+\sum_{j \neq i} Y_j})/(1 + \frac{x^*}{1-p_i})$. Let us prove that $\gamma_i \geq \frac{2}{3}$ whenever $z_i^* > 0$. Indeed, Jensen's inequality implies $\gamma_i \geq 2/[(1 + \sum_{j \neq i} y_j)(1 + \frac{x^*}{1-p_i})]$, and since

$$\sum_{j \neq i} y_j = \sum_{j \neq i} \frac{2z_j^*}{1 + x^*/(1-p_j)} \leq \frac{2}{1+x^*} \sum_{j \neq i} z_j^* = \frac{2(1-x^*-z_i^*)}{1+x^*},$$

we deduce

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

When $x^*=0$ the inequality $\gamma_i \geq \frac{2}{3}$ follows at once. For the case $x^*>0$ we note that, since (z^*, x^*) is a basic solution, exactly one of the two inequalities involving each z_i^* must be tight so that $z_i^* > 0$ implies $z_i^* = \frac{p_i}{2}(1 + \frac{x^*}{1-p_i})$ which plugged into the previous bound yields

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

The conclusion follows since the latter is at least $\frac{2}{3}$ for all $x^* \in [0, 1]$. \square

Note that if we have three clients with characteristics $p_1 = \varepsilon$, $p_2 = p_3 = 1 - \varepsilon/2$, $v_1 = 1$, $v_2 = v_3 = \varepsilon/(2 - \varepsilon)$, we have that $V_{LP_2} = V^* = \varepsilon$ and $A_4^* = \{1, 2, 3\}$, whose value approaches $5\varepsilon/6$ for ε small. Thus our algorithm is not better than a $5/6$ -approximation. On the other hand, our analysis in the last proposition is tight. Indeed, the client-by-client bound γ_i that we use to prove the result can actually achieve the value $2/3$. To see this note that in the same three-client instance above the value z_1 in LP_2 is $\varepsilon/2$, while in $V(A_4^*)$ the coefficient accompanying v_1 approaches $\varepsilon/3$.

5 A PTAS for the single item case

In this section we provide polynomial time approximation scheme (PTAS) for the single item case, namely, given $\varepsilon > 0$ we show how to compute (in time which is polynomial for fixed ε) a strategy that yields a value within a factor $(1+\varepsilon)$ from the optimum. Recall that after introducing (P) in §2 we noted that this problem has an equivalent combinatorial version, namely

$$\max_{A \subseteq N} V_A := \sum_{R \subseteq A} v_R p_R \tag{7}$$

where $p_R = \prod_{i \in R} p_i \cdot \prod_{i \in A \setminus R} (1 - p_i)$. We say a client $i \in N$ is *large* if $p_i > \varepsilon$ and *small* otherwise. We will show that the following algorithm computes an $(1+O(\sqrt{\varepsilon}))$ -approximate solution

ALGORITHM — PTAS
for all $A_\ell \subseteq N$ with $|A_\ell| \leq \frac{1}{\varepsilon} \ln(\frac{1}{\varepsilon})$ **do**
 → Discard all large clients in $N \setminus A_\ell$
 → **for all** $k = 1, \dots, n$ compute $V_{A_\ell \cup \{1, \dots, k\}}$
Output the best solution found.

The running time of the algorithm can be estimated as follows. We try at most $\binom{n}{\frac{1}{\varepsilon} \ln(1/\varepsilon)}$ subsets A_ℓ and, for each one, we evaluate $V_{A_\ell \cup \{1, \dots, k\}}$ for $k = 1, \dots, n$. Since each evaluation takes $O(n^3)$, the overall running time of the algorithm is $O\left(n^{4 + \frac{\ln(1/\varepsilon)}{\varepsilon}}\right)$, which is polynomial for fixed ε .

In order to establish the $(1+O(\sqrt{\varepsilon}))$ approximation guarantee we proceed as follows. First we show that restricting the optimization in (7) to strategies with at most $\frac{1}{\varepsilon} \ln(1/\varepsilon)$ large clients, accounts for a loss factor of at most $(1+O(\varepsilon))$. Therefore, it suffices to check all strategies A which contain a subset of large clients A_ℓ of size at most $\frac{1}{\varepsilon} \ln(1/\varepsilon)$. Secondly, for the small clients we will show that the best threshold strategy is near-optimal. This is precisely what the algorithm does.

5.1 Restricting the optimal strategy

We show that there is a near optimal solution for (P) whose sum of probabilities is not too large.

Lemma 6 *For small $\varepsilon > 0$ there is $A \subseteq N$ with $\sum_{i \in A} p_i \leq \ln(1/\varepsilon)$ such that $V^* \leq (1+3\varepsilon)V_A$.*

Proof. Take A^* optimal for (P) and assume without loss of generality that $A^* = \{1, \dots, n\}$. If $\sum_{i \in A^*} p_i \leq \ln(1/\varepsilon)$ just take $A = A^*$. Otherwise, let $A := \{1, \dots, k\} \subseteq A^*$ with k the largest integer satisfying $\sum_{i=1}^k p_i \leq \ln(1/\varepsilon)$. The probability that no customer in A accepts can be bounded as

$$q := \mathbb{P}[\text{no customer in } A \text{ accepts}] = \prod_{i \in A} (1 - p_i) \leq e^{-\sum_{i=1}^k p_i} \leq e^{-(\ln(1/\varepsilon)-1)} = e\varepsilon$$

where the first inequality follows from $1 - x \leq e^{-x}$, and the second from the definition of k .

When no customer in A accepts the strategy A gets zero revenue while A^* gets at most v_k . On the other hand, if some customers in A accept then strategy A gets more revenue than A^* , and of course revenue at least v_k . Thus, for $\varepsilon \leq (3 - e)/3e$, we have $V_{A^*} - V_A \leq v_k q \leq \frac{q}{1-q} V_A \leq 3\varepsilon V_A$. \square

According to this Lemma, we may restrict to strategies A such that $\sum_{i \in A} p_i \leq \ln(1/\varepsilon)$, which accounts for a loss of at most a factor $1 + O(\varepsilon)$.

5.2 Modifying the instance

The next step is to consider a modified instance which is easier to solve and from which we recover an approximate solution for the original problem (P) . The idea is to *split* each small client $p_i \leq \varepsilon$ into several *clones* with an even smaller probability δ , and to show that this does not affect too much the expected revenue. More precisely, given an instance $I = \{(v_1, p_1), \dots, (v_n, p_n)\}$ with set of clients N and a small $\delta > 0$, we build an instance I' which contains all large customers in I and also, for each small customer $i \in I$ we add to I' a number q_i of clones with value v_i and probability δ . Here q_i is chosen to be the largest integer satisfying

$$1 - p_i \leq (1 - \delta)^{q_i}. \quad (8)$$

Observe that the quantity $\tilde{p}_i := 1 - (1 - \delta)^{q_i}$ can be made arbitrarily close to p_i by taking δ small enough. For convenience we denote \tilde{I} the perturbed instance of I in which each small p_i is changed into \tilde{p}_i . By continuity, for any $A \subset N$ the value V_A in both instances I and \tilde{I} are arbitrarily close. Hence, by choosing δ small enough we can make the optimal values of I and \tilde{I} to lie within a factor arbitrarily close to 1.

Let N' denote the set of clients in I' and let $\sigma : N' \rightarrow N$ so that for each $i \in N$ the set $\sigma^{-1}(i)$ denotes the corresponding clones in N' (just one if client i is large). We claim that I' and \tilde{I} have very similar optimal values. Roughly speaking, we interpret the value of a strategy A in I as the value under a *modified random process* in which a client $i \in A$ accepts if and only if at least one clone in $\sigma^{-1}(i)$ accepts in I' . As usual, for every $j \in N'$ we let Y'_j denote a Bernoulli random variable with $\mathbb{P}(Y'_j = 1)$ equal to p_j if j is a large, and equal to δ for all clones j of small customers. We will also denote $Y_i = \max\{Y'_j : j \in \sigma^{-1}(i)\}$ for each $i \in N$ and we observe that these Y_i 's are independent Bernoulli's of parameter p_i for $i \in N$ large and \tilde{p}_i for $i \in N$ small. Using these variables the value $V_{A'}$ of a strategy $A' \subset N'$ in the instance I' can be expressed as

$$V_{A'} = \mathbb{E} \left[\frac{\sum_{j \in A'} v_j Y'_j}{\sum_{j \in A'} Y'_j} \right]$$

while the value \tilde{V}_A of a strategy $A \subset N$ in the instance \tilde{I} is

$$\tilde{V}_A = \mathbb{E} \left[\frac{\sum_{i \in A} v_i Y_i}{\sum_{i \in A} Y_i} \right].$$

We will show that both A and $A' = \sigma^{-1}(A)$ yield the same revenue with high probability. To this end we note that the probability that all customers in $\sigma^{-1}(i)$ reject the offer in I' is

$$\mathbb{P}[\sigma^{-1}(i) \text{ reject}] = (1 - \delta)^{q_i} = 1 - \tilde{p}_i, \quad (9)$$

while the the next lemma shows that the probability that more than one clone in $\sigma^{-1}(i)$ accepts simultaneously is very small.

Lemma 7 *Let R' be the respondent customers in $A' = \sigma^{-1}(A)$. If $\varepsilon < \frac{1}{2}$ then for each small client $i \in N$ we have $\mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 2] \leq 2\tilde{p}_i^2 \leq 2p_i^2$.*

Proof. We observe that

$$\begin{aligned} \mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 2] + \mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 1] &= 2\mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 2] + \mathbb{P}[|\sigma^{-1}(i) \cap R'| = 1] \\ &\leq \mathbb{E}[|\sigma^{-1}(i) \cap R'|] = q_i \delta \leq \ln \frac{1}{1 - \tilde{p}_i} \leq \frac{\tilde{p}_i}{1 - \tilde{p}_i}, \end{aligned}$$

where the last inequalities follow by taking logarithm in (9). Now, using (9) once again we have

$$\mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 1] = 1 - \mathbb{P}[\sigma^{-1}(i) \text{ reject}] = \tilde{p}_i,$$

which combined to the previous estimate yields $\mathbb{P}[|\sigma^{-1}(i) \cap R'| \geq 2] \leq \tilde{p}_i / (1 - \tilde{p}_i) - \tilde{p}_i \leq 2\tilde{p}_i^2$. \square

We will say that a (random) respondent set R' for the instance I' is *good* if $|\sigma^{-1}(i) \cap R'| \leq 1$ for all small clients $i \in N$. Otherwise we will say that R' is *bad*.

Lemma 8 *Let $A \subseteq N$ with $\sum_{i \in A} \tilde{p}_i \leq \ln(1/\varepsilon)$ and $A' = \sigma^{-1}(A)$. Then $V_{A'} \geq (1 + O(\sqrt{\varepsilon}))\tilde{V}_A$.*

Proof. First notice that $\tilde{V}_A \leq (1 + O(\sqrt{\varepsilon})) \sum_{R:|R| \leq 1/\sqrt{\varepsilon}} v_R \tilde{p}_R$. Indeed,

$$\begin{aligned} \tilde{V}_A &= \sum_{i \in A} v_i \left(\sum_{R \ni i; |R| \leq 1/\sqrt{\varepsilon}} \frac{\tilde{p}_R}{|R|} + \sum_{R \ni i; |R| > 1/\sqrt{\varepsilon}} \frac{\tilde{p}_R}{|R|} \right) \leq \sum_{i \in A} v_i \left(\sum_{R \ni i; |R| \leq 1/\sqrt{\varepsilon}} \frac{\tilde{p}_R}{|R|} + \sum_{R \ni i} \tilde{p}_R \sqrt{\varepsilon} \right) \\ &\leq \sum_{i \in A} v_i \sum_{R \ni i; |R| \leq 1/\sqrt{\varepsilon}} \frac{\tilde{p}_R}{|R|} + \sum_{i \in A} v_i \tilde{p}_i \sqrt{\varepsilon} \leq \sum_{R:|R| \leq 1/\sqrt{\varepsilon}} v_R \tilde{p}_R + (1 - \ln \varepsilon) \sqrt{\varepsilon} \tilde{V}_A \end{aligned}$$

where the last inequality comes from

$$\tilde{V}_A = \sum_{i \in A} v_i \tilde{p}_i \mathbb{E}\left(\frac{1}{1 + S_i}\right) \geq \sum_{i \in A} v_i \tilde{p}_i \frac{1}{1 + \sum_{i \in A} \tilde{p}_i} \geq \sum_{i \in A} v_i \tilde{p}_i \frac{1}{1 - \ln \varepsilon}.$$

On the other hand, $V_{A'}$ can be bounded as follows:

$$V_{A'} = \sum_{R \subseteq A} \sum_{R': \sigma(R') = R} v_{R'} p_{R'} \geq \sum_{R \subseteq A; |R| \leq 1/\sqrt{\varepsilon}} \left(\sum_{R' \text{ good}; \sigma(R') = R} v_{R'} p_{R'} \right). \quad (10)$$

Now, for all R' good we have $v_{R'} = v_R$, while $\sum_{R' \text{ good}; \sigma(R') = R} p_{R'}$ is the probability that a random R' is good and $\sigma(R') = R$, that is to say

$$\prod_{i \notin R} \mathbb{P}[|\sigma^{-1}(i) \cap R'| = 0] \prod_{i \in R} \mathbb{P}[|\sigma^{-1}(i) \cap R'| = 1] \geq \prod_{i \notin R} (1 - \tilde{p}_i) \prod_{i \in R} (\tilde{p}_i - 2\tilde{p}_i^2) = \tilde{p}_R \prod_{i \in R} (1 - 2\tilde{p}_i),$$

where the inequality follows from Lemma 7. Plugging the latter into (10) we get

$$V_{A'} \geq \sum_{R \subseteq A; |R| \leq 1/\sqrt{\varepsilon}} v_R \tilde{p}_R \prod_{i \in R} (1 - 2\tilde{p}_i) \geq \sum_{R \subseteq A; |R| \leq 1/\sqrt{\varepsilon}} v_R \tilde{p}_R (1 - 2\varepsilon)^{1/\sqrt{\varepsilon}}.$$

The lemma follows by combining this bound with that obtained for \tilde{V}_A . \square

Lemma 9 Let $A \subseteq N$ and set $A' = \sigma^{-1}(A)$. Then $V_{A'} \leq (1 + O(\varepsilon))\tilde{V}_A$.

Proof. Note that from the definition of $Y_i = \max\{Y'_j : j \in \sigma^{-1}(i)\} \leq \sum_{j \in \sigma^{-1}(i)} Y'_j$. Then

$$\begin{aligned} V_{A'} &= \sum_{i \in A} \sum_{j \in \sigma^{-1}(i)} v_j p_j \mathbb{E} \left(\frac{1}{1 + \sum_{k \neq j} Y'_k} \right) \leq \sum_{i \in A} v_i \sum_{j \in \sigma^{-1}(i)} p_j \mathbb{E} \left(\frac{1}{1 + \sum_{\ell \in A, \ell \neq i} \sum_{k \in \sigma^{-1}(\ell)} Y'_k} \right) \\ &\leq \sum_{i \in A} v_i \mathbb{E} \left(\frac{1}{1 + \sum_{\ell \neq i} Y_\ell} \right) \sum_{j \in \sigma^{-1}(i)} p_j. \end{aligned}$$

To conclude observe that if i is big, then $\sigma^{-1}(i)$ is a singleton and $\sum_{j \in \sigma^{-1}(i)} p_j = p_i$, while for i small we have $\sum_{j \in \sigma^{-1}(i)} p_j = \delta q_i \leq (1 + O(\varepsilon))\tilde{p}_i$. \square

5.3 Proof of approximation guarantee

Theorem 10 For small $\varepsilon > 0$ our algorithm computes a $(1 + O(\sqrt{\varepsilon}))$ -approximate solution in polynomial time.

Proof. Let V' be the optimal value of instance I' . Using Lemma 6 there is A' with $\sum_{i \in A'} p_i \leq \ln \frac{1}{\varepsilon}$ whose value is within a factor $(1 + O(\varepsilon))$ of V' . Let A_ℓ denote the set of large customers in A' so that we necessarily have $|A_\ell| \leq \frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}$. Now, let also A_s be the set of small customers in A' , so that $A_\ell \cup A_s = A'$. Note that we can assume without loss of generality that A_s contains all clones of the highest valued small clients, except eventually for a threshold small client i^* for which only a strict subset of its clones $\sigma^{-1}(i^*)$ is considered. Indeed, since all small clients have the same probability δ , if $j \notin A_s$ and $k \in A_s$ are two small clients with $v_j > v_k$, then exchanging j and k in A_s increases the value of A' .

We observe that strategy $A = \sigma(A') \setminus \{i^*\}$ is of the form considered in the algorithm, so that denoting A^* the best strategy found by the algorithm we have

$$V_A \leq V_{A^*} \leq V^*. \quad (11)$$

Now, consider the strategy $A'' = \sigma^{-1}(A)$. The only difference between A'' and A' is that A'' does not contain any clone in $\sigma^{-1}(i^*)$. Thus the value of A'' is smaller than that of A' only if no customer in A'' accepts and some customer in A' does, in which case the difference is exactly v_{i^*} . This event occurs with probability at most ε , from where we get $V_{A'} \leq V_{A''} + \varepsilon v_{i^*}$. Moreover, if $A'' \neq A'$, then $V_{A'} \geq v_{i^*}$, as otherwise including one more clone j of customer i^* to A' would lead to a strategy whose value is at least $(1 - \delta)V_{A'} + \delta v_{i^*}$. This implies $(1 - \varepsilon)V_{A'} \leq V_{A''}$, which combined with Lemmas 8 and 9 yields

$$\tilde{V}_A = (1 + O(\sqrt{\varepsilon}))V_{A''} \geq (1 + O(\sqrt{\varepsilon}))V_{A'} = (1 + O(\sqrt{\varepsilon}))V'. \quad (12)$$

Clearly, by a continuity argument, we can fix a priori δ small enough so that the \tilde{p}_i 's are all sufficiently close to the p_i 's in order to guarantee $V_B = (1 + O(\varepsilon))\tilde{V}_B$ for all $B \subseteq N$. In particular such a δ guarantees that $V_A = (1 + O(\varepsilon))\tilde{V}_A$, which combined with (11) and (12) leads to

$$V_{A^*} \geq V_A \geq (1 + O(\sqrt{\varepsilon}))V'. \quad (13)$$

Now, by Lemma 6 there is a set $\bar{A} \subseteq N$, whose sum of probabilities is at most $\ln(1/\varepsilon)$, such that $V^* \leq (1 + O(\varepsilon))V_{\bar{A}}$. Again, the choice of δ implies $V_{\bar{A}} \leq (1 + O(\varepsilon))\tilde{V}_{\bar{A}}$, while Lemma 8 gives

$$\tilde{V}_{\bar{A}} \leq (1 + O(\sqrt{\varepsilon}))V_{\sigma^{-1}(\bar{A})} \leq (1 + O(\sqrt{\varepsilon}))V',$$

so that $V^* \leq (1 + O(\sqrt{\varepsilon}))V'$. The latter together with (13) concludes the proof. \square

6 Final Remarks

Throughout the paper we have assumed that customers are selected according to a first-come-first served discipline. We have argued that this policy is reasonable in our short-term setting, since it only assumes that customers response times are independent and identically distributed, and moreover, this seems to be the only alternative as it may be too difficult, or too late, to discriminate among clients. In addition, note that our approximation results concerning threshold policies in §3, use as comparison benchmark (upper bound) \bar{V} , the solution of a linear program which is also an upper bound for the revenue of *any* rationing policy, so that the potential gain of using a more sophisticated policy is still at most $O(1/\sqrt{m})$. In particular, for 10 items, our simple policies are guaranteed to obtain at least 80% of the revenue of any rationing policy, and this fraction grows to 1 as the number of items to be sold grows.

Let us now comment on further research directions. First, a natural extension, related to stochastic knapsack [6], is the following. As before, customers accept the offer with probabilities p_i and contribute a value v_i if they buy an item. The difference is that customer $i \in C$ demands $s_i \geq 1$ units. The function $V(y)$ becomes more complicated and the problem is now NP-hard: for $p_i = 1$ we have the standard knapsack problem. Furthermore, in this situation computing $V(y)$ is actually #P-complete [9]. An interesting question is whether the techniques derived in this paper may be adapted to design approximation algorithms for this more general problem.

Another possible extension is to consider a multistage problem as in revenue management. In this situation the short-term assumption is no longer valid and one may address offers in several rounds, maybe a fixed number k . One would now look for an adaptive online algorithm that, given the current situation and the number of rounds left, computes the optimal subset of clients to whom the offer should be addressed in the next round.

A third direction, related to the airline-industry application mentioned in §1, concerns pricing. When clients can be contacted in a personalized fashion by email, the airline may quote a different price for each selected client. Naturally, the clients' probability of accepting an offer will decrease as the price increases, while the value of a client increases with the selling price. Therefore, a natural model is to associate with client i a decreasing function $p_i : \mathbb{R}_+ \rightarrow [0, 1]$ and an increasing function $v_i : \mathbb{R}_+ \rightarrow \mathbb{R}_+$, representing the acceptance probability and value for a given price respectively. One may argue about the actual shape of these functions, for instance, it is natural to take $v_i(\cdot)$ as an affine function. The problem is now to select the customers to whom the offer will be made available together with a price for each selected customer, so as to maximize the expected revenue.

7 Appendix: Numerical experiments

In this section we present some statistics that illustrate the numerical performance of the algorithms A_1^* through A_4^* discussed previously. We also compare these solutions with two additional heuristics that work well in practice. The first heuristic starts from $A = \phi$ and recursively adds or removes a single element to get a maximal increase in value, namely

HEURISTIC 5 — IN-OUT

initialize $A_5^* = \phi$

find B with $|B \Delta A_5^*| = 1$ and V_B maximal

if $V_B \leq V_{A_5^*}$ *then stop, otherwise set* $A_5^* = B$ *and repeat.*

The resulting solution may be suboptimal, as for $m = 1$ and $n = 3$ with $p_1 = 0.5$ $p_2 = 0.9$ $p_3 = 1$ $v_1 = 75$ $v_2 = 50$ $v_3 = 47$, we have $A_5^* = \{1, 3\}$ while the optimal solution is $A^* = \{1, 2\}$. This example reveals another unexpected feature: if $A^k \subseteq C$ denotes an optimal solution for the original problem with the additional constraint $|A| \leq k$, then a customer included in A^k need not

be considered when we relax to $|A| \leq k+1$. As counter-intuitive as it may seem, this occurs in the example above where $A^1 = \{3\}$ while $A^2 = A^3 = A^* = \{1, 2\}$.

The second heuristic builds on the previous idea considering not only the addition/removal of elements but also the swapping of an element in the current solution with an element from outside

HEURISTIC 6 — SWAP

initialize $A_6^* = \phi$

find B of the form $B = (A_6^* \setminus \{i\}) \cup \{j\}$ with V_B maximal

if $V_B \leq V_{A_6^*}$ then stop, otherwise set $A_6^* \leftarrow B$ and repeat.

The resulting solution A_6^* is still suboptimal as for $m = 1$ and $n = 5$ with

Probabilities	(p_i)	0.7	0.6	0.7	0.2	1.0
Values	(v_i)	94	69	68	67	64

where $A_6^* = \{1, 5\}$ while the optimal solution is $A^* = \{1, 2, 3\}$.

We run a series of 10000 random instances for $m = 1$ and $n = 5$ (with p_i 's and v_i 's generated uniformly in $[0, 1]$ and $[0, 100]$ respectively), which we solved exactly as well as by each of the 6 algorithms. The table below shows the percentage of problems solved to optimality by each method, the minimal and average optimality ratios attained over the 10000 trials, and the average running time per instance (in milliseconds).

Algorithm		% opt	Min ratio	Avg ratio	Time
<i>optimal</i>	(A^*)	100.0	1.0000	1.0000	12.9
MAX- k	(A_1^*)	98.2	0.9590	0.9999	3.2
ADD- m	(A_2^*)	77.4	0.7531	0.9928	0.6
MAX-AVG	(A_3^*)	51.2	0.7358	0.9793	0.7
LP-RELAX	(A_4^*)	88.3	0.9451	0.9989	18.3
IN-OUT	(A_5^*)	99.8	0.9489	0.9999	3.9
SWAP	(A_6^*)	100.0	0.9972	1.0000	6.9

Table 1. Algorithm performance: $m = 1$, $n = 5$, 10000 instances

The lower bound \tilde{V} was always in the range $[0.5, 0.86]$ with an average of 0.66, while the upper bound \bar{V} was in $[1.00, 1.49]$ with 1.18 as average. The exercise was repeated with 200 instances of size $m = 1$ and $n = 10$, with the following results.

Algorithm		% opt	Min ratio	Avg ratio	Time
<i>optimal</i>	(A^*)	100.0	1.0000	1.0000	1611.2
MAX- k	(A_1^*)	93.0	0.9916	0.9998	23.5
ADD- m	(A_2^*)	47.5	0.8168	0.9771	0.4
MAX-AVG	(A_3^*)	11.5	0.8169	0.9564	2.3
LP-RELAX	(A_4^*)	74.0	0.9775	0.9988	21.5
IN-OUT	(A_5^*)	99.0	0.9918	0.9999	15.4
SWAP	(A_6^*)	100.0	1.0000	1.0000	36.5

Table 2. Algorithm performance: $m = 1$, $n = 10$, 200 instances

We note the excellent performance attained by MAX- k and LP-RELAX, at a small fraction of the execution time required by the brute force method. Despite the fact that ADD- m and MAX-AVG are dominated by MAX- k which finds the best threshold solution, these methods are still close to the optimum on average and are much faster so they could be useful in applications where speed is an issue. On average ADD- m outperforms MAX-AVG both in optimal value and execution time, with

the additional advantage that ADD- m applies for all m and not only for $m = 1$ as MAX-AVG. The IN-OUT and SWAP heuristics get closer to the optimal value with a moderate increase in execution time. The situation is similar for $m = 3$, $n = 10$ as shown below (note that MAX-AVG and LP-RELAX do not apply for $m > 1$):

Algorithm	% opt	Min ratio	Avg ratio	Time
<i>optimal</i> (A^*)	100.0	1.0000	1.0000	1601.8
MAX- k (A_1^*)	88.0	0.9813	0.9995	23.5
ADD- m (A_2^*)	72.5	0.9049	0.9968	5.0
IN-OUT (A_5^*)	97.0	0.9738	0.9995	66.9
SWAP (A_6^*)	100.0	1.0000	1.0000	202.2

Table 3. Algorithm performance: $m = 3$, $n = 10$, 200 instances

When we increase the dimension of the problem the exact solution takes too long to be computed by brute force and all the running times grow, with an increasing relative advantage for ADD- m . For $m = 15$ and $n = 50$ we get the following values and running times

Algorithm	MAX- k	ADD- m	IN-OUT	SWAP
Value	1057.9	1057.9	1058.2	1058.2
Time	3265	140	41740	494321

Note that the IN-OUT and SWAP running times increase significantly. We do not know if these heuristics run in polynomial time or not.

Acknowledgments. We thank two anonymous referees and the associate editor for many very helpful comments and suggestions.

References

- [1] Ball M., Queyranne M., *Toward robust revenue management: Competitive analysis of online booking*, Operations Research 57, 950-963 (2009).
- [2] Bult J.R., Wansbeek T. *Optimal selection for direct mail*, Marketing Science 14, 378-394 (1995).
- [3] Chriqui C., Robillard P., *Common bus lines*, Transportation Science 9, 115-121 (1975).
- [4] Cominetti R., Vaisman J., *A sharp uniform bound for the distribution of a sum of Bernoulli random variables*, Preprint, Departamento de Ingeniería Matemática, U. de Chile, (2008).
- [5] Cormen T., Leiserson C., Rivest R., Stein C. Introduction to Algorithms, MIT Press, Cambridge (2001).
- [6] Dean B., Goemans M.X., Vondrak J., *Approximating the stochastic knapsack problem: The benefit of adaptivity*, Mathematics of Operations Research 33, 945-964 (2008).
- [7] Jogdeo K., Samuels S.M., *Monotone convergence of binomial probabilities and a generalization of Ramanujan's equation*, The Annals of Mathematical Statistics 39, 1191-1195 (1968).
- [8] Littlewood, K. *Forecasting and Control of Passengers*, 12th AGIFORS Symposium Proc., 95-128 (1972).
- [9] Platzman L., Ammnos J., Bartholdi III J., *A simple and efficient algorithm to compute tail probabilities from transforms*, Operations Research 36, 137-144 (1988).
- [10] Talluri K.T., Van Ryzin G.J., The Theory and Practice of Revenue Management, Springer-Verlag, New York (2005).
- [11] Siegel A., *Median bounds and their applications*, Journal of Algorithms 38, 184-236 (2001).
- [12] Vaisman J., *Convergencia fuerte del método de medias sucesivas para operadores lineales no-expansivos*, Memoria de Ingeniería Civil Matemática, Universidad de Chile, 2005.