More Revenue from Two Samples via Factor Revealing SDPs

More Revenue from Two Samples via Factor Revealing SDPs

CONSTANTINOS DASKALAKIS, Massachusetts Institute of Technology
MANOLIS ZAMPETAKIS, Massachusetts Institute of Technology

We consider the classical problem of selling a single item to a single bidder whose value for the item is drawn from a regular distribution F, in a "data-poor" regime where F is not known to the seller, and very few samples from F are available. Prior work [9] has shown that one sample from F can be used to attain a $1/2$-factor approximation to the optimal revenue, but it has been challenging to improve this guarantee when more samples from F are provided, even when two samples from F are provided. In this case, the best approximation known to date is 0.509, achieved by the Empirical Revenue Maximizing (ERM) mechanism [2]. We improve this guarantee to 0.558, and provide a lower bound of 0.65. Our results are based on a general framework, based on factor-revealing Semidefinite Programming relaxations aiming to capture as tight as possible a superset of product measures of regular distributions, the challenge being that neither regularity constraints nor product measures are convex constraints. The framework is general and can be applied in more abstract settings to evaluate the performance of a policy chosen using independent samples from a distribution and applied on a fresh sample from that same distribution.

CCS Concepts: • **Theory of computation → Algorithmic game theory and mechanism design**; **Algorithmic mechanism design**.

Additional Key Words and Phrases: revenue maximization, two samples, mechanism design, semidefinite relaxation

ACM Reference Format:
Constantinos Daskalakis and Manolis Zampetakis. 2020. More Revenue from Two Samples via Factor Revealing SDPs. In *Proceedings of the 21st ACM Conference on Economics and Computation (EC '20), July 13–17, 2020, Virtual Event, Hungary.* ACM, New York, NY, USA, 16 pages. https://doi.org/10.1145/3391403.3399543

1 INTRODUCTION

Consider a seller who is selling an item to a single buyer, whose value for the item is drawn from some distribution F. The seller knows F, but does not know the realized value $v \sim F$ of the buyer, which is known to the buyer. What is the best sales protocol for the seller to use in this scenario? A fundamental result in mechanism design states that among all sales protocols that the seller could use, a rather simple one is optimal, namely posting a well-chosen price x^* and letting the buyer decide whether they are willing to pay x^* to buy the item. Of course, the optimal price to use is any $x^* \in \arg\max\{x \cdot (1 - F(x))\}$, interpreting F as a cumulative density function. This result was obtain by Riley and Zeckhauser [21] and by Myerson [19], who also extended it to the multi-buyer case, and it justifies mathematically centuries-old selling practices.

The question that arises, however, is how the seller learns F. The standard justification used in the theory of Bayesian Mechanism Design is that the seller has sold the same type of item before to buyers from the same population, and has estimated F from her interaction with these buyers via econometric analysis of their purchasing behavior. Alternatively, the seller might have done

EC '20, July 13–17, 2020, Virtual Event, Hungary

ACM ISBN 978-1-4503-7975-5/20/07…$15.00
https://doi.org/10.1145/3391403.3399543

market research to estimate F. Regardless, chances are that the seller does not know F exactly, and one must be cautious about the potential loss in revenue resulting from the errors in estimating F. In particular, one would like to avoid "over-fitting" their mechanism to the estimated distribution $\hat{F}$.

The afore-described discrepancy between the standard Bayesian Mechanism Design model, which assumes distributional knowledge, and practical reality has motivated lots of work in Econometrics and, more recently, Algorithmic Game Theory, where most of the focus has been on learning good mechanisms given *sample access* to the underlying distribution [1, 3, 4, 6–10, 10, 11, 14–18, 22]. This work has essentially settled how many samples are needed to learn approximately optimal mechanisms in the single-item single-bidder setting discussed earlier, and has made lots of progress in a variety of single-item multi-bidder settings as well as several more general settings.

In particular, prior work has focused on the *data-rich* regime, where the goal is to characterize the number of samples that are necessary to identify a $(1-\varepsilon)$-optimal mechanism. It is typically shown that, as $\varepsilon \longrightarrow 0$, the sample complexity scales polynomially in $\frac{1}{\varepsilon}$ and other parameters of the setting. For example, in the single-item single-bidder setting discussed earlier, it has been shown [16] that $\tilde{\Theta}(\frac{1}{\varepsilon^3})$-samples are necessary and sufficient to identify a mechanism whose revenue is $(1-\varepsilon)\cdot OPT$, when the distribution F is regular.[1]

The focus of prior work notwithstanding, this paper studies instead the *data-poor* regime. We are interested in whether reasonably good mechanisms can be identified when one has a small number of samples from the underlying distribution. As it turns out, this data-poor regime is quite important in applications. Interestingly, it is quite important in the design of online ad auctions, despite the fact that ad exchanges run millions of auctions per day. The reason is that these auctions involve different combinations of website, viewer, advertiser, etc., leading to an exponential explosion of combinations of features that influence the value of the advertiser. Ultimately, it is quite common that the ad exchange has very few observations from past auctions whose features are similar enough to those of the auction that it is about to run, and it needs to use these past few observations intelligently to turn the knobs of the auction that it will run in the next instance, e.g. compute reserve prices for the advertisers, in order to increase the resulting revenue.

So, what is known about learning good mechanisms in the data-poor regime? Not much, even in the single-bidder single-item setting described above. An elegant geometric argument by Dhangwatnotai et al. [9] shows that, if one is given one sample $x \sim F$ from a regular distribution F, and uses sample x as a take-it-or-leave-it price against a new buyer $y \sim F$ from the same distribution, the resulting expected revenue is at least $\frac{1}{2}\cdot$ OPT. Specifically:

$$\mathbb{E}_{x\sim F}\left[\mathbb{E}_{y\sim F}[x\cdot \mathbb{1}_{y\geq x}]\right] \geq \frac{1}{2}\cdot \max\{p\cdot(1-F(p))\}.$$

The LHS calculates the revenue from using price equal to sample x. Indeed, we receive revenue equal to the first sample as long as the second sample is larger than the first. The RHS is precisely $\frac{1}{2}\cdot$ OPT, given our discussion in the beginning of this section. More recently, Huang et al. [16] showed that the $\frac{1}{2}$-approximation guarantee cannot be improved upon by any deterministic mechanism that is chosen based on the value of a single sample, while Fu et al. [12] showed that the guarantee can be slightly improved using randomization.

Given that a $\frac{1}{2}$-factor approximation is the best that is achievable by deterministic mechanisms given one sample from a regular distribution F, can we do better given, say, two samples from

[1]Regularity, defined formally in Section 2, is a tail condition on distributions that is standard in Bayesian Mechanism Design literature. In the context of learning good mechanisms from samples, regularity allows multiplicative approximations to the optimal revenue. Indeed, it is easy to see that without any tail restriction on the distribution it is impossible to attain multiplicative approximations, by considering distributions that are 0 with overwhelming probability, and take some large value with some small probability.

the distribution? As it turns out, analyzing the revenue of mechanisms chosen using two samples from a distribution has been challenging analytically, without any progress until a very recent paper of Babaioff et al [2]. In this work, they consider the revenue of the natural *Empirical Revenue Maximizing (ERM)* mechanism. ERM takes two samples $x, y \sim F$, and compares the maximum M to the minimum m of the two samples. If $M \geq 2m$, the item is priced M, otherwise the item is priced m. Babaioff et al show that ERM with two samples achieves expected revenue at least $0.509 \cdot \text{OPT}$. Specifically:

$$\mathbb{E}_{x,y\sim F}\left[\mathbb{E}_{z\sim F}[\text{Revenue}_{ERM(x,y)}(z)]\right] \geq 0.509 \cdot \max\{p \cdot (1 - F(p))\}.$$

Analyzing the revenue of ERM from two samples was challenging and tedious in the work of [2], and it appears difficult to improve thosee guarantees or show that they cannot be improved. Motivated by this apparent difficulty, in this paper we target developing a general framework, based on Semidefinite Programming, for analyzing the revenue attainable by mechanisms chosen using independent samples from a distribution F and evaluated on a fresh independent sample from that same distribution. Our framework is general and can be applied in more abstract settings to evaluate the performance of a policy chosen using independent samples from a distribution and applied on a fresh sample from that same distribution, however we present it in the context of analyzing the revenue of ERM from two samples for simplicity. Applying our framework in this context, we are able to show Theorem 5.1, which states that a rounded version of ERM from two samples attains revenue $\sim 0.56 \cdot \text{OPT}$, namely

$$\mathbb{E}_{x,y\sim F}\left[\mathbb{E}_{z\sim F}[\text{Revenue}_{\overline{ERM}(x,y)}(z)]\right] \geq 0.558 \cdot \max\{p \cdot (1 - F(p))\}.$$

Moreover, our framework provides an associated lower bound, Theorem 6.1, which states that there exists a regular distribution F under which ERM from two samples cannot do better than $\sim 0.642 \cdot \text{OPT}$.

The framework is based on constructing a factor-revealing Semidefinite Programming relaxation, whose optimal value bounds the approximation factor of a rounded version of ERM. The starting point for developing the framework is the realization that one way to bound the approximation factor of ERM is to write a mathematical program that finds the worst-case distribution. The variables in this program are the probabilities assigned by the distribution on its (discretized) support, the objective function is a degree 3 polynomial in the distribution, and the regularity constraints are expressible as inequalities that degree 2 polynomials of the distribution must satisfy. As such, solving this mathematical program is computationally challenging. Our framework proposes to relax this program and optimize instead over a three-dimensional tensor that expresses the joint distribution of the three samples. By optimizing over joint distributions we, of course, provide more power to the adversary so this is a valid relaxation. The benefit is that both the objective function and the regularity constraints are now linear in the tensor! The only question is how to restrict the power of the adversary enough so that the relaxation is not too loose. We impose symmetry constraints on the tensor and require that its two-dimensional slices are positive semidefinite. We also cut the tail of the distribution and discretize its support appropriately, and argue that these operations we perform are not real restrictions to the adversary in coming up with a worst-case distribution. More precisely, we characterize how to account for these operations in our relaxation. Our relaxation steps are explained in Section 3. Ultimately we arrive at a number of SDPs that we solve to calculate the worst-case approximation factor for ERM. The implementation details are discussed in Section 4 and the results of solving the SDPs and how to combine them to get our result are discussed in Section 5. Section 6 shows our lower bound.

2 PRELIMINARIES

Notation. We use small bold letters $\boldsymbol{x}$ to refer to real vectors in finite dimension $\mathbb{R}^d$ and capital bold letters $\boldsymbol{A}$ to refer to matrices in $\mathbb{R}^{d\times\ell}$. Similarly, a function with image in $\mathbb{R}^d$ is represented by a small and bold letter $\boldsymbol{f}$. We use $\boldsymbol{I}_d$ to refer to the identity matrix in d dimensions. We may skip the subscript when the dimensions are clear. We use $\boldsymbol{E}_{i,j}$ to refer to the all zero matrix with one 1 at the (i, j) entry. Let $\boldsymbol{A} \in \mathbb{R}^{d\times d}$, we define $\boldsymbol{A}^\flat \in \mathbb{R}^{d^2}$ to be the standard vectorization of $\boldsymbol{A}$. Let also $\mathcal{S}_d$ be the set of all the symmetric $d \times d$ matrices and $\mathcal{Q}_d$ the cone of the positive semidefinite matrices. $\mathbb{E}[x]$ is the expected value of the random variable x and $\mathrm{Var}[x]$ is the variance of x.

2.1 Single Buyer - Single Item Pricing

We consider the setting where the seller has a single item to sell to a single buyer. The buyer has a private valuation $v \in \mathbb{R}_+$ for this item which we assume is a random variable distributed with respect to some distribution $\mathcal{F}$, with support $\mathbb{R}_+$, cumulative distribution function F and probability density function f. We assume that the seller has access to exactly *two* samples $\hat{v}_1, \hat{v}_2$ from the distribution $\mathcal{F}$ and otherwise $\mathcal{F}$ is unknown to the seller. The goal of the seller based on $\hat{v}_1$ and $\hat{v}_2$ is to compute a price p and set it as a reserve price for the item. We assume that the buyer has quasi-linear utility that is equal to $u(p) = v - p$ if she decides to buy the item and $v = 0$ otherwise.

Regular Distributions and Revenue Curve. In this work we make the standard assumption that the distribution $\mathcal{F}$ is *regular*. The easiest way to explain this assumption is in terms of the *revenue curve* of $\mathcal{F}$. Let $q \in [0, 1]$, that represents the quantile of the distribution $\mathcal{F}$, we define $r_{\mathcal{F}}(q)$, the value of the revenue curve at p, that is equal to $r_{\mathcal{F}}(q) = (1 - q)F^{-1}(q)$. The regularity assumption, is equivalent to the condition that the revenue curve $r(q)$ is a concave function of q. Let $\mathcal{R}$ be the set of all regular distributions over $\mathbb{R}_+$.

Optimal Auction and Expected Revenue. For any fixed price p and given that the valuation of the buyer is distributed according to $\mathcal{F}$, the expected revenue of the seller if she sets the reserve price p is equal to $R_{\mathcal{F}}(p) = p(1 - F(p))$. Based on the celebrated result of Myerson [19], over the set of all the possible auctions the one that maximizes the revenue of the seller corresponds to the optimal reserved price auction. For a regular distribution $\mathcal{F}$ this optimal price is equal to $p^\star_{\mathcal{F}} = \operatorname{argmin} p(1 - F(p))$ and the optimal expected revenue is equal to $\mathrm{OPT}_{\mathcal{F}} = R_{\mathcal{F}}(p^\star_{\mathcal{F}})$.

Empirical Revenue Maximization. We follow the notation of [2] and we define the Empirical Revenue Maximization mechanism (ERM) to be the mechanism that computes the optimal price assuming that the unknown distribution $\mathcal{F}$ is equal to the empirical distribution $\hat{\mathcal{F}}$ of the samples. In particular when two samples are only observed the exact form of ERM is the following

$$\mathsf{ERM}(\hat{v}_1, \hat{v}_2) = \begin{cases} \min(\hat{v}_1, \hat{v}_2) \text{ if } \max(\hat{v}_1, \hat{v}_2) \leq 2 \cdot \min(\hat{v}_1, \hat{v}_2) \\ \max(\hat{v}_1, \hat{v}_2) \text{ otherwise} \end{cases}. \tag{2.1}$$

The corresponding expected revenue is equal to

$$\textsc{Rev}_{\mathcal{F}}(\mathsf{ERM}) = \mathop{\mathbb{E}}_{x\sim\mathcal{F}}\left[\mathop{\mathbb{E}}_{y\sim\mathcal{F}}[R_{\mathcal{F}}(\mathsf{ERM}(x, y))]\right]. \tag{2.2}$$

Rounded Empirical Revenue Maximization. In this paper instead of analyzing the vanilla ERM mechanism we analyze a rounded version of ERM for which we are able to theoretically bound its performance using the semidefinite relaxation that we introduce in the next section. Let $\rho > 1$ be the *rounding parameter* and G_ρ be a gridding of the positive real line $\mathbb{R}_+$ with the following points $G_\rho = \{\rho^i \mid i \in \mathbb{Z}\}$. For every positive real number x we define $g_\rho(x)$ to be the largest number in G_ρ that is smaller than x, that is $g_\rho(x) = \max_{t \in G_\rho : t \leq x} t$. Then the rounded empirical revenue maximization mechanism is the following

$$\overline{\mathsf{ERM}}_\rho(\hat{v}_1, \hat{v}_2) = \mathsf{ERM}(g_\rho(\hat{v}_1), g_\rho(\hat{v}_2)). \tag{2.3}$$

The corresponding expected revenue is equal to

$$\textsc{Rev}_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho\right) = \mathop{\mathbb{E}}_{x \sim \mathcal{F}}\left[\mathop{\mathbb{E}}_{y \sim \mathcal{F}}\left[R_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho(x, y)\right)\right]\right]. \tag{2.4}$$

Approximation Ratio. The approximation ratios that we are interested in are the following

$$\alpha^\star = \min_{\mathcal{F} \in \mathcal{R}} \frac{\textsc{Rev}_{\mathcal{F}}(\mathsf{ERM})}{\mathrm{OPT}_{\mathcal{F}}}, \ \alpha_\rho^\star = \min_{\mathcal{F} \in \mathcal{R}} \frac{\textsc{Rev}_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho\right)}{\mathrm{OPT}_{\mathcal{F}}}. \tag{2.5}$$

3 SEMIDEFINITE PROGRAMMING RELAXATION

We start with a formulation of the problem that we are trying to solve as an optimization problem in the space of all distributions. Then we explain how we can apply a discretization that leads to an optimization problem in a finite dimensional space. Finally we relax this problem to get a positive semidefinite problem that can be approximately solved and most importantly it admits a certification of optimality due to the existence of a dual program. Our strategy is to construct a program that abstractly has the following form

$$\begin{aligned} \alpha_\rho^\star = \min_{\mathcal{F}, \mathrm{OPT}, p^\star} \ & \frac{1}{\mathrm{OPT}} \text{ (revenue of } \overline{\mathsf{ERM}}_\rho \text{ for distribution } \mathcal{F}) \\ \text{s.t. } & (\mathcal{F} \text{ satisfies the regularity constraints}) \\ & \text{(the revenue for any price is at most OPT)} \\ & \text{(the revenue for price } p^\star \text{ is exactly OPT)} \end{aligned}$$

where the variables are: the distribution $\mathcal{F}$, the value of the optimal possible revenue OPT, and the optimal price $p^\star$. Then we add the appropriate constraints make sure that $\mathcal{F}$ is regular and that the optimum is realized at price $p^\star$ and has value OPT. We start with the definition of the objective function and then we move to the definition of the regularity and the optimality constraints. Then we describe how to relax this constraints to get ultimately a finite dimensional optimization problem and then relax this more to get a positive semidefinite program.

3.1 Objective Function

Let f be the probability density function of $\mathcal{F}$, from the definition of $\alpha_\rho^\star$ we have that the objective $b(\mathcal{F})$ of our program is equal to

$$b(\mathcal{F}) = \mathop{\mathbb{E}}_{x \sim \mathcal{F}}\left[\mathop{\mathbb{E}}_{y \sim \mathcal{F}}\left[R_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho(x, y)\right)\right]\right] = \frac{1}{\mathrm{OPT}_{\mathcal{F}}} \int_{\mathbb{R}_+} \int_{\mathbb{R}_+} R_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho\,(x, y)\right) \cdot f(x) \cdot f(y) \cdot dxdy.$$

Now we observe that for every $x \in [\rho^i, \rho^{i+1}]$, the value of $\overline{\mathsf{ERM}}_\rho(x, y)$ does not change, and the same is true for y. Hence we can define the following quantities

$$f_i = F\left(\rho^{i+1}\right) - F\left(\rho^i\right), \ \overline{\mathsf{ERM}}_\rho^{(i,j)} = \overline{\mathsf{ERM}}_\rho(\rho^i, \rho^j) \tag{3.1}$$

Therefore we have the following

$$b(\mathcal{F}) = \frac{1}{\mathrm{OPT}} \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} R_{\mathcal{F}}\left(\overline{\mathsf{ERM}}_\rho^{(i,j)}\right) \cdot f_i \cdot f_j$$

and from the definition of the function $R_{\mathcal{F}}(\cdot)$ we have that

$$b(\mathcal{F}) = \frac{1}{\mathrm{OPT}} \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \overline{\mathrm{ERM}}_{\rho}^{(i,j)} \left(1 - F(\overline{\mathrm{ERM}}_{\rho}^{(i,j)})\right) \cdot f_i \cdot f_j$$
$$= \frac{1}{\mathrm{OPT}_{\mathcal{F}}} \sum_{i \in \mathbb{Z}} \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} \overline{\mathrm{ERM}}_{\rho}^{(i,j)} \mathbb{1}\left(k \geq \overline{\mathrm{ERM}}_{\rho}^{(i,j)}\right) \cdot f_i \cdot f_j \cdot f_k \tag{3.2}$$

3.2 Regularity Constraints

We start with the regularity constraints, namely that $\mathcal{F} \in \mathcal{R}$. The regularity constraints are concavity constraints on the revenue curve $r_{\mathcal{F}}$. There are a lot of different ways to express concavity constraints and we use the following that is also shown in the Figure 1, for any $x, y \in [0, 1]$ and any $z \in [x, y]$, the point $(z, r_{\mathcal{F}}(z))$ is above the line that connects $(x, r_{\mathcal{F}}(x))$ and $(y, r_{\mathcal{F}}(y))$.

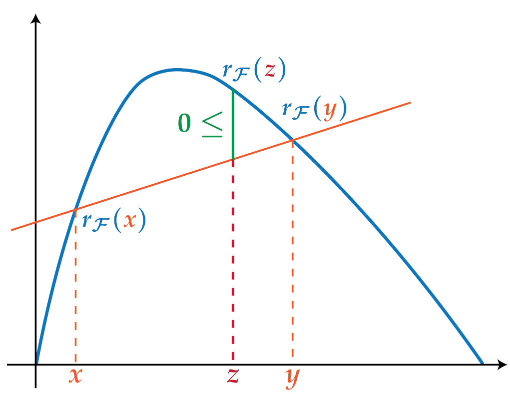


Fig. 1. Illustration of the regularity conditions.

More precisely,

$$\frac{r_{\mathcal{F}}(x) - r_{\mathcal{F}}(y)}{x - y}(z - y) + r_{\mathcal{F}}(y) \leq r_{\mathcal{F}}(z) \; \forall x, y \in [0, 1], z \in [x, y].$$

These constraints are expressed in the quantile space of the distributions $\mathcal{F}$. But our final goal is to formulate a problem in the space $\mathbb{R}_+$ of prices. By the definition of the revenue curve it is easy to see that the regularity constraints are equivalent with the following

$$F(z)(1 - F(x))x - F(z)(1 - F(y))y - F(y)(1 - F(x))x + F(y)(1 - F(z))z$$
$$\geq F(x)(1 - F(z))z - F(x)(1 - F(y))y. \; \forall x, y, z \in \mathbb{R}_+ \tag{3.3}$$

3.3 Optimality Constraints

We now describe the constraints that ensure that the values of the parameters OPT and $p^{\star}$ are correct. First we have that the revenue at any price should be at most OPT and which is equivalent with the following

$$x(1 - F(x)) \leq \mathrm{OPT} \; \forall x \in \mathbb{R}_+. \tag{3.4}$$

Also we need to ensure that the value OPT is achievable at the point $p^{\star}$ or

$$p^{\star}(1 - F(p^{\star})) \geq \mathrm{OPT}. \tag{3.5}$$

So using (3.2), (3.3), (3.4), (3.5) we have the following equation.

$$\alpha^\star_\rho = \min_{\mathcal{F},p^\star,\mathrm{OPT}} \frac{1}{\mathrm{OPT}} \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \overline{\mathrm{ERM}}_\rho^{(i,j)} \mathbb{1}\left(k \geq \overline{\mathrm{ERM}}_\rho^{(i,j)}\right) \cdot f_i \cdot f_j \cdot f_k$$
$$\text{s.t. } F(z)(1-F(x))x - F(z)(1-F(y))y - F(y)(1-F(x))x + F(y)(1-F(z))z$$
$$\geq F(x)(1-F(z))z - F(x)(1-F(y))y.\ \forall x,y,z \in \mathbb{R}_+$$
$$x(1-F(x)) \leq \mathrm{OPT}\ \forall x \in \mathbb{R}_+$$
$$p^\star(1-F(p^\star)) \geq \mathrm{OPT}. \tag{3.6}$$

3.4 Relaxing Regularity and Optimality Constraints

The first step towards constructing a finite dimensional relaxation of (3.6) is to relax the infinitely many regularity constraints. The idea is to keep only constraints that involve numbers in $\mathbb{R}_+$ that are of the form ρ^i with $i \in \mathbb{Z}$. So for regularity we keep only the constraints that have $x = \rho^i$, $y = \rho^j$ and $z = \rho^k$ and doing simple calculations we get the following

$$\sum_{k'=-\infty}^{k} \sum_{i'=i+1}^{\infty} f_{k'} f_{i'} \rho^i - \sum_{k'=-\infty}^{k} \sum_{j'=j+1}^{\infty} f_{k'} f_{j'} \rho^j - \sum_{j'=-\infty}^{j} \sum_{i'=i+1}^{\infty} f_{j'} f_{i'} \rho^i + \sum_{j'=-\infty}^{j} \sum_{k'=k+1}^{\infty} f_{j'} f_{k'} \rho^k$$
$$\geq \sum_{i'=-\infty}^{i} \sum_{k'=k+1}^{\infty} f_{i'} f_{k'} \rho^k - \sum_{i'=-\infty}^{i} \sum_{j'=j+1}^{\infty} f_{i'} f_{j'} \rho^j\ \forall i,j,k. \in \mathbb{Z} \tag{3.7}$$

Using the same simplification for optimality constraints we get the following

$$\rho^i \sum_{i'=i+1}^{\infty} f_{i'} \leq \mathrm{OPT}\ \forall i \in \mathbb{Z}. \tag{3.8}$$

$$\rho^{i^\star} \sum_{i'=i^\star+1}^{\infty} f_{i'} \geq \frac{\mathrm{OPT}}{\rho}. \tag{3.9}$$

Observe that in the last optimality constraint we forced $p^\star = \rho^{i^\star}$ which is not necessarily true but we also relaxed the right hand side from OPT to $\frac{\mathrm{OPT}}{\rho}$. If $p^\star \in [\rho^{i^\star}, \rho^{i^{\star+1}}]$ then we know that for any $x \in [\rho^{i^\star}, \rho^{i^{\star+1}}]$, $(1-F(x)) \leq (1-F(\rho^{i^\star}))$ and also $x \leq \rho^{i^\star} \cdot \rho$. From this observation we get that $R_\mathcal{F}(x) \leq \rho \cdot R_\mathcal{F}(\rho^{i^\star})$ and this holds for any distribution $\mathcal{F}$. Hence $R_\mathcal{F}(p^\star) \leq \rho \cdot R_\mathcal{F}(\rho^{i^\star})$ and therefore (3.9) is indeed a relaxation of the condition (3.5). Also because of the slack that we have now introduced between (3.8) and (3.9) we assume without loss of generality that $\mathrm{OPT} = \rho^{i_\mathrm{OPT}}$, with $i_\mathrm{OPT} \in \mathbb{Z}$.

We conclude therefore that we can replace the optimization over continuous distributions $\mathcal{F}$ in (3.6), with optimization over sequences $(f_i)_{i\in\mathbb{Z}}$ such that $f_i \geq 0$ and $\sum_{i\in\mathbb{Z}} f_i = 1$. Also observe in (3.6) the optimization with respect to $p^\star$ and OPT does not lead to a convex objective and for this reason we keep the corresponding parameters $i^\star$ and i_OPT in the relaxed version of the optimization program, as hyperparameters. Hence we get the following relaxation

$$
\begin{aligned}
\alpha_\rho(i^\star, i_{\mathrm{OPT}}) = \min_{(f_i)_{i\in\mathbb{Z}}} \ & \frac{1}{\rho^{i_{\mathrm{OPT}}}} \sum_{i\in\mathbb{Z}} \sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}} \overline{\mathsf{ERM}}_\rho^{(i,j)} \mathbb{1}\left(k \geq \overline{\mathsf{ERM}}_\rho^{(i,j)}\right) \cdot f_i \cdot f_j \cdot f_k \\
\text{s.t.} \ & \sum_{k'=-\infty}^{k} \sum_{i'=i+1}^{\infty} f_{k'} f_{i'} \rho^i - \sum_{k'=-\infty}^{k} \sum_{j'=j+1}^{\infty} f_{k'} f_{j'} \rho^j - \sum_{j'=-\infty}^{j} \sum_{i'=i+1}^{\infty} f_{j'} f_{i'} \rho^i + \sum_{j'=-\infty}^{j} \sum_{k'=k+1}^{\infty} f_{j'} f_{k'} \rho^k \\
& \geq \sum_{i'=-\infty}^{i} \sum_{k'=k+1}^{\infty} f_{i'} f_{k'} \rho^k - \sum_{i'=-\infty}^{i} \sum_{j'=j+1}^{\infty} f_{i'} f_{j'} \rho^j \ \forall i, j, k \in \mathbb{Z} \\
& \rho^i \sum_{i'=i+1}^{\infty} f_{i'} \leq \rho^{i_{\mathrm{OPT}}} \ \forall i \in \mathbb{Z} \\
& \rho^{i^\star} \sum_{i'=i^\star+1}^{\infty} f_{i'} \geq \rho^{i_{\mathrm{OPT}}-1} \\
& \sum_{i\in\mathbb{Z}} f_i = 1 \\
& f_i \geq 0 \ \forall i \in \mathbb{Z}.
\end{aligned}
\tag{3.10}
$$

Where because of what we have discussed already, we know that $\alpha_\rho^\star \geq \min_{i^\star, i_{\mathrm{OPT}}} \alpha_\rho(i^\star, i_{\mathrm{OPT}})$.

3.5 Relaxing to a Finite Dimensional Program

In the previous section we show how to move from the space of all continuous distributions to a relaxation in the space of sequences $(f_i)_{i\in\mathbb{Z}}$. Now we relax our program more to get a relaxation in a finite dimensional space. To do this we have to cut the upper and the lower tails of the sequence $(f_i)_{i\in\mathbb{Z}}$. For this we will use another hyperparameter N that together with ρ measures the quality of the discretization.

For the lower tail we define the variable $F = \sum_{i=-\infty}^{i_{\mathrm{OPT}}-N} f_i$ and we drop all the regularity conditions for $i, j, k \leq i_{\mathrm{OPT}} - N$. We also drop all the optimality conditions for $i \leq i_{\mathrm{OPT}} - N$ but observe that this is not actually a relaxation since for $j \leq i_{\mathrm{OPT}}$ the revenue $R_{\mathcal{F}}(\rho^j)$ is always less than $\mathrm{OPT} = \rho^{i_{\mathrm{OPT}}}$. To make sure that the program remains a relaxation, we replace in our objective every term that contains an index that is less that $i_{\mathrm{OPT}} - N$ with 0.

For the upper tail we assume that $f_j = 0$ for $j > i_{\mathrm{OPT}} + N$. With this assumption we only loose an additive term of $\left(\frac{1}{\rho}\right)^N$ from the approximation ratio, as we will see in Lemma 3.1. Therefore we get the following finite dimensional relaxation of 3.10, where when we write $\sum_{i=-\infty}^{j} f_i$ we mean the sum $\sum_{i=i_{\mathrm{OPT}-N}}^{j} f_i + F$. The variables of this program now are $(f_i)_{i\in[i_{\mathrm{OPT}}-N,\, i_{\mathrm{OPT}}+N]}$, F. Let also for simplicitly $m_- = i_{\mathrm{OPT}} - N$ and $m_+ = i_{\mathrm{OPT}} + N$.

$$\alpha_{\rho,N}(i^\star, i_{\mathrm{OPT}}) = \min_{(f_i),F} \frac{1}{\rho^{i_{\mathrm{OPT}}}} \sum_{i=m_-}^{m_+} \sum_{j=m_-}^{m_+} \sum_{k=m_-}^{m_+} \overline{\mathrm{ERM}}_\rho^{(i,j)} \mathbb{1}\left(k \geq \overline{\mathrm{ERM}}_\rho^{(i,j)}\right) \cdot f_i \cdot f_j \cdot f_k$$

$$\text{s.t.} \sum_{k'=-\infty}^{k} \sum_{i'=i+1}^{m_+} f_{k'} f_{i'} \rho^i - \sum_{k'=-\infty}^{k} \sum_{j'=j+1}^{m_+} f_{k'} f_{j'} \rho^j - \sum_{j'=-\infty}^{j} \sum_{i'=i+1}^{m_+} f_{j'} f_{i'} \rho^i + \sum_{j'=-\infty}^{j} \sum_{k'=k+1}^{m_+} f_{j'} f_{k'} \rho^k$$

$$\geq \sum_{i'=-\infty}^{i} \sum_{k'=k+1}^{m_+} f_{i'} f_{k'} \rho^k - \sum_{i'=-\infty}^{i} \sum_{j'=j+1}^{m_+} f_{i'} f_{j'} \rho^j \; \forall i,j,k \in \mathbb{Z}$$

$$\rho^i \sum_{i'=i+1}^{m_+} f_{i'} \leq \rho^{i_{\mathrm{OPT}}} \; \forall i \in [m_-, m_+]$$

$$\rho^{i^\star} \sum_{i'=i^\star+1}^{m_+} f_{i'} \geq \rho^{i_{\mathrm{OPT}}-1} \tag{3.11}$$

$$\sum_{i=m_-}^{m_+} f_i + F = 1$$

$$f_i \geq 0 \; \forall i \in [m_-, m_+]$$

$$F \geq 0.$$

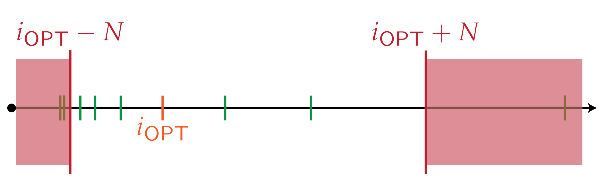


Fig. 2. Illustration of the cut of the tails of the distribution.

Based on the definitions of $\alpha_\rho(i^\star, i_{\mathrm{OPT}})$ and $\alpha_{\rho,N}(i^\star, i_{\mathrm{OPT}})$ in (3.10) and (3.11) respectively we can prove the following lemma.

Lemma 3.1. *It holds that* $\alpha_{\rho,N}(i^\star, i_{\mathrm{OPT}}) \geq \left(1 - \frac{1}{\rho^N}\right) \alpha_\rho(i^\star, i_{\mathrm{OPT}})$.

Proof. The proof follows immediately from Lemma 47 of [5]. □

3.6 SDP Relaxation

The main issue with the program (3.11) is that it involves an optimization over product measures which is in general intractable. To make (3.11) tractable we consider the SDP relation by considering as variables the general tensor $f_{i,j,k}$[2]. We define also the matrix $\boldsymbol{F}_i = (f_{i,j,k})_{(j,k)}$ and for simplicity we set $n = 2N$, and we translate the indices from the interval $[i_{\mathrm{OPT}} - N, i_{\mathrm{OPT}} + N]$ to the interval $[1, n]$. Observe that with this indexing of the variables the i_{OPT} will always be equal to $n/2$ and hence we shouldn't treat it as a hyperparameter of the problem anymore. Since we have a general three dimensional tensor now we include also some symmetry constraints that trivially hold when

[2]For simplicity we assume that this tensor contains the variable F too.

the tensor is a product measure. We also add the positive semidefinite constraints $\boldsymbol{F}_i \succeq 0$ which again are trivially satisfied by any product measure. Then our final relaxation is the following.

$$\begin{aligned}
\bar{\alpha}_{\rho,n}(i^\star) = \min_{(f_i),F}\ & \frac{1}{\rho^{i_{\mathrm{OPT}}}} \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} \overline{\mathsf{ERM}}_\rho^{(i,j)} \mathbb{1}\left(k \geq \overline{\mathsf{ERM}}_\rho^{(i,j)}\right) \cdot f_{i,j,k} \\
\text{s.t.} & \sum_{k'=-\infty}^{k}\sum_{i'=i+1}^{m_+} f_{k',i',\ell}\rho^i - \sum_{k'=-\infty}^{k}\sum_{j'=j+1}^{m_+} f_{k',j',\ell}\rho^j - \sum_{j'=-\infty}^{j}\sum_{i'=i+1}^{m_+} f_{j',i',\ell}\rho^i + \sum_{j'=-\infty}^{j}\sum_{k'=k+1}^{m_+} f_{j',k',\ell}\rho^k \\
& \geq \sum_{i'=-\infty}^{i}\sum_{k'=k+1}^{m_+} f_{i',k',\ell}\rho^k - \sum_{i'=-\infty}^{i}\sum_{j'=j+1}^{m_+} f_{i',j',\ell}\rho^j \ \forall i,j,k,\ell \in [1,n] \\
& \rho^i \sum_{i'=i+1}^{m_+} f_{i',j,k} \leq \rho^{i_{\mathrm{OPT}}} \ \forall i,j,k \in [1,n] \\
& \rho^{i^\star} \sum_{i'=i^\star+1}^{m_+} f_{i',j,k} \geq \rho^{n/2-1} \ \forall j,k \in [1,n] \\
& \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} f_{i,j,k} = 1 \qquad (3.12) \\
& \boldsymbol{F}_i \succeq 0 \ \forall i \in [1,n] \\
& f_{i,j,k} \geq 0 \ \forall i,j,k \in [1,n] \\
& f_{i,j,k} = f_{j,i,k} \ \forall i,j,k \in [1,n] \\
& f_{i,j,k} = f_{i,k,j} \ \forall i,j,k \in [1,n]
\end{aligned}$$

Based on the arguments that we have explained in this section and using the Lemma 3.1 the following Theorem follows.

Theorem 3.2. *It holds that* $\alpha_\rho^\star \geq \left(1 - \frac{1}{\rho^{n/2}}\right) \min_{i^\star \in [n/2,n]} \left\{\bar{\alpha}_{\rho,n}(i^\star)\right\}$.

4 TECHNICAL DETAILS

As we have already explained our goal is to find a primal-dual pair for the program (3.12) and then use Theorem 3.2 to prove our main result. For this purpose, we will use the solver SCS from [20], in the interface of cvxpy. Before diving into some optimization on (3.12) that were necessary, we explain how we can get an accurate result from the guarantees of the SCS solver. Without loss of generality we restrict the program (3.12) such that $\rho^{i_{\text{OPT}}+n/2} = 1$.

4.1 Approximation Error from Approximate Solutions of (3.12)

In order to simplify the notation and be more close with the notation used in [20] we assume that the program (3.12) can be written in the following form.

$$\begin{aligned} &\min_{\boldsymbol{x}} \boldsymbol{c}^T \boldsymbol{x} \\ &s.t.\ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{s} = \boldsymbol{b} \\ &(\boldsymbol{x}, \boldsymbol{s}) \in \mathcal{K} \end{aligned}$$

The dual of the above program is

$$\begin{aligned} &\max_{\boldsymbol{y}}\ -\boldsymbol{b}^T \boldsymbol{y} \\ &s.t.\ -\boldsymbol{A}^T \boldsymbol{y} + \boldsymbol{r} = \boldsymbol{c} \\ &(\boldsymbol{y}, \boldsymbol{r}) \in \mathcal{K}^* \end{aligned}$$

From the termination criteria in Section 3.5 of [20], we have that the SCS solver will return a primal-dual pair $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{s}}, \hat{\boldsymbol{r}}, \hat{\boldsymbol{y}})$ such that

$$\|\boldsymbol{A}\hat{\boldsymbol{x}} + \hat{\boldsymbol{s}} - \boldsymbol{b}\|_2 \le \varepsilon_p(1 + \|\boldsymbol{b}\|_2), \tag{4.1}$$

$$\left\|-\boldsymbol{A}^T\hat{\boldsymbol{y}} + \hat{\boldsymbol{r}} - \boldsymbol{c}\right\|_2 \le \varepsilon_d(1 + \|\boldsymbol{c}\|_2), \tag{4.2}$$

$$\left\|\boldsymbol{c}^T\hat{\boldsymbol{x}} + \boldsymbol{b}^T\hat{\boldsymbol{y}}\right\|_2 \le \varepsilon_g(1 + \left|\boldsymbol{c}^T\hat{\boldsymbol{x}}\right| + \left|\boldsymbol{b}^T\hat{\boldsymbol{y}}\right|) \tag{4.3}$$

where the ε_p, ε_d and ε_g are the primal residual, the dual residual and the duality gap tolerances respectively. We have also the guarantee that $\hat{\boldsymbol{x}} \in \mathcal{K}$ and $\hat{\boldsymbol{y}} \in \mathcal{K}^*$. For all the experiment that we ran the most important parameter is ε_d, the other two are at least 10^{-4} smaller, so we set $\varepsilon = \varepsilon_d$. Also for all the experiments that we ran we had $\|\boldsymbol{b}\| = 1$, $\|\boldsymbol{c}\| \le 0.02$ and obviously by the construction of our program we have that $\left|\boldsymbol{c}^T\boldsymbol{x}\right| \le 1$ and $\left|\boldsymbol{b}^T\boldsymbol{y}\right| \le 1$. Hence we have that

$$\|\boldsymbol{A}\hat{\boldsymbol{x}} + \hat{\boldsymbol{s}} - \boldsymbol{b}\|_2 \le 2 \cdot 10^{-4}\varepsilon, \tag{4.4}$$

$$\left\|-\boldsymbol{A}^T\hat{\boldsymbol{y}} + \hat{\boldsymbol{r}} - \boldsymbol{c}\right\|_2 \le 1.02 \cdot \varepsilon, \tag{4.5}$$

$$\left\|\boldsymbol{c}^T\hat{\boldsymbol{x}} + \boldsymbol{b}^T\hat{\boldsymbol{y}}\right\|_2 \le 3 \cdot 10^{-4}\varepsilon. \tag{4.6}$$

Now we can take the weighted average of the primal constraints with respect to the values of the dual and we conclude that for any feasible point $\boldsymbol{x}$ it holds that

$$-\boldsymbol{b}^T\hat{\boldsymbol{y}} \le \boldsymbol{c}^T\hat{\boldsymbol{x}} + (-\boldsymbol{A}^T\hat{\boldsymbol{y}} + \hat{\boldsymbol{r}} - \boldsymbol{c})^T\hat{\boldsymbol{x}}.$$

But from the constraints of our primal problem (3.12) we have that $\boldsymbol{x}$ is a distribution, so $\|\boldsymbol{x}\|_1 = 1$ which implies $\|\boldsymbol{x}\|_2 \le 1$, hence from the guarantees of the SCS solver we have that

$$-\boldsymbol{b}^T\hat{\boldsymbol{y}} \le \boldsymbol{c}^T\hat{\boldsymbol{x}} + 1.02 \cdot \varepsilon$$

from which we get the following lemma

Lemma 4.1. *We assume that* $\|\boldsymbol{b}\|_2 \leq 1$, $\|\boldsymbol{c}\|_2 \leq 0.02$, $\varepsilon_p \leq 10^{-4}\varepsilon$ *and* $\varepsilon_g \leq 10^{-4}\varepsilon$. *If* $(\hat{\boldsymbol{x}}, \hat{\boldsymbol{s}}, \hat{\boldsymbol{r}}, \hat{\boldsymbol{y}})$ *is a primal-dual solution returned by the* SCS *solver on the program* (3.12) *and under the aforementioned assumptions we have that*

$$\boldsymbol{b}^T\hat{\boldsymbol{y}} - 1.02\varepsilon \leq \bar{\alpha}_{\rho,n}(i^\star).$$

4.2 Reducing the Number of Variables

One of the biggest challenges towards running the program (3.12) is that the number of variables increases very fast as n^3. Even a constant reduction of this number is very important in the ability to run bigger programs. For this reason we observe that because of the symmetry constraints that we have in the end of (3.12) we don't need exactly n^3 free variables. Instead for any variable $f_{i,j,k}$ we need only one representative up to permutations of the indices (i, j, k). This reduces the number of variables from n^3 to $\binom{n+3}{3}$ which significantly improves the performance of the SCS solver.

5 RESULTS FROM THE SCS SOLVER

For our final result we choose the parameters $\rho = 1.08$, $n = 100$ and we solve all the programs $\bar{\alpha}_{1.08,100}(i)$ for all $i \in [49, 99]$ and $\varepsilon = 0.01$. The minimum value that we get is larger than 0.581 for $i = 87$. Hence by Lemma 4.1 and by the fact that this was the minimum result we have that

$$\min_{i \in [49,99]} \bar{\alpha}_{1.08,100}(i) \geq 0.581 - 1.01 \cdot 0.01 = 0.5709.$$

Finally from Theorem 3.2 and using the above inequality we get that

$$\alpha^\star_{1.08} \geq \left(1 - \left(\frac{1}{1.08}\right)^{50}\right) \min_{i \in [49,99]} \bar{\alpha}_{1.08,100}(i) \geq 0.978 \cdot 0.5709 \geq 0.558$$

from which our main theorem follows.

Theorem 5.1. *There exists a mechanism that has access to two samples from a regular distribution* $\mathcal{F}$*, and achieves revenue at least* $0.558 \cdot \mathrm{OPT}$.

The whole proof of Theorem 5.1 consists of the code together with the dual certificates for all the SDP's $\bar{\alpha}_{1.08,100}(i)$, where $i \in [49, 99]$[3].

We next present the primal values that we get for the case $\rho = 1.08$ and $i^\star = 87$. Since this primal solution is huge, we present the marginal distribution that we observe with respect to the first index. More precisely, we consider the distribution

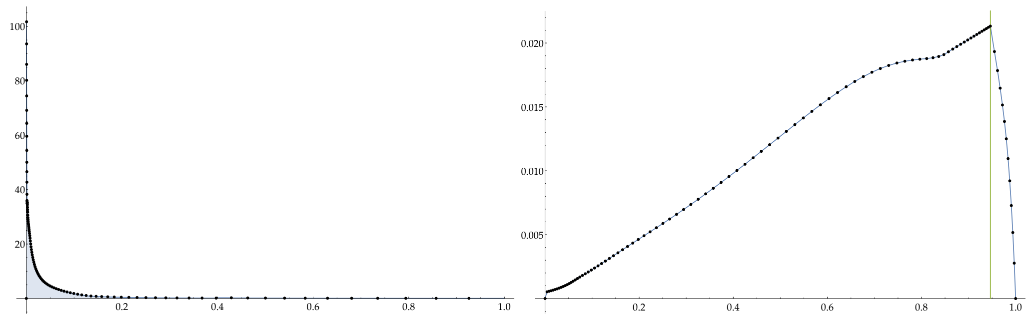
$$f_i = \sum_{j,k} x_{i,j,k}. \tag{5.1}$$

As we can see in the next two plots the solution that we get is almost regular and achieves the maximum exactly at the value that we were expecting. From the optimality constraints in the program (3.12) it is clear that the value of the optimal revenue of $\mathcal{F}$ should be $\rho^{i_{\mathrm{OPT}}-1}$ which is exactly what we observe in the experiments.

6 LOWER BOUND OF APPROXIMATION RATION OF ERM

In this section we prove a non-trivial lower bound on the value $\alpha^\star$. To build the counter-example that is used in our lower bound we used our SDP relaxation program (3.12), but we also verified the validity of the lower bound as we show in this section. The result that we want to prove is the following.

[3]All this information can be found in the following anonymous github repository: https://github.com/anonymousAuthorPrime/Two-Samples-SDP.



(a) Piecewise constant probability density function. (b) Revenue curve.

Fig. 3. Properties of the marginal $\mathcal{F}$ defined in (5.1) of the approximate solution of $\bar{\alpha}_{1.08,100}(87)$ solved using SCS.

Theorem 6.1. *It holds that $\alpha^\star \leq 0.642$.*

Proof of Theorem 6.1. In order to prove this theorem we will describe a regular distribution that achieves this approximation ratio. To define the distribution and make sure that it is a regular distribution we start with $n = 16$ of the form $(q, r_{\mathcal{F}}(q))$ on the revenue curve of the distribution. Then we define a complete revenue curve by taking the linear interpolation of these points and we can easily check that the result is a concave curve. One thing that we also get directly from this description is the maximum revenue and the optimal reserve price since both of them will correspond to one of the initial points on the revenue curve. The points (q_i, r_i) that we start with are presented in the following table together with the value of the variables $c_i^{(1)}$ and $c_i^{(2)}$ that are computed according to the following equations and will be used to define the probability density function of the distribution for our counter-example.

$$c_i^{(1)} = \frac{r_{i+1} - r_i}{q_{i+1} - q_i}, \; c_i^{(2)} = r_i - c_i^{(1)} \cdot s_i \tag{6.1}$$

Then our distribution can be define as the following piecewise function

$$f(x) = \sum_{i=1}^{15} \mathbb{1}\left\{x \in \left[\frac{i-1}{n-1}, \frac{i}{n-1}\right]\right\} \cdot \frac{c_i^{(1)} + c_i^{(2)}}{\left(c_i^{(1)} + x\right)^2}. \tag{6.2}$$

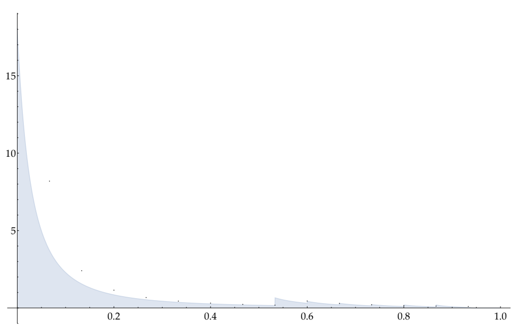
Using the definition of the density function we can run numeric integration to get an upper and a lower bound estimate for the following integral, that measures the revenue achieved by the ERM with two samples mechanism when under the distribution with density f. To simplify the expression we will use the notation $m_{x,y} = \min(x, y)$ and $M_{x,y} = \max(x, y)$.

$$\begin{aligned}
\bar{\ell} &= \int_0^1 \int_0^1 R_{\mathcal{F}}(\mathsf{ERM}(x, y)) \cdot f(x) \cdot f(y) \cdot dxdy = \\
&= \int_0^1 \int_0^1 \int_0^1 \mathbb{1}\{z \geq m_{x,y}\} \cdot \mathbb{1}\{2m_{x,y} \geq M_{x,y}\} \cdot f(x) \cdot f(y) \cdot f(z) \cdot dx \cdot dy \cdot dz + \\
&\quad + \int_0^1 \int_0^1 \int_0^1 \mathbb{1}\{z \geq M_{x,y}\} \cdot \mathbb{1}\{M_{x,y} > 2m_{x,y}\} \cdot f(x) \cdot f(y) \cdot f(z) \cdot dx \cdot dy \cdot dz
\end{aligned}$$

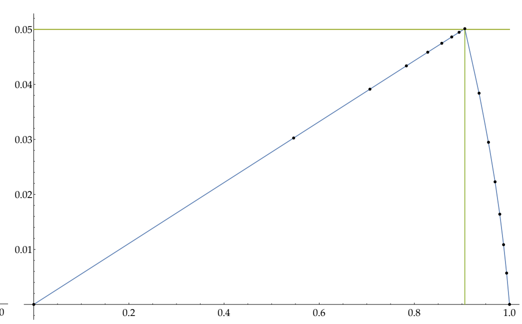
To numerically compute an accurate value for the above integral we use the *Global Adaptive* strategy of `mathematica`, where at every step the algorithm chooses to split the subinterval that

i	q_i	r_i	$c_i^{(1)}$	$c_i^{(2)}$
1	0	0	0.0554	0
2	0.5462	0.0303	0.0554	0
3	0.7066	0.0391	0.0553	0
4	0.7832	0.0434	0.0552	0.0001
5	0.8281	0.0458	0.0551	0.0002
6	0.8576	0.0475	0.0550	0.0003
7	0.8785	0.0486	0.0546	0.0006
8	0.8940	0.0495	0.0546	0.0006
9	0.9060	0.0501	-0.3910	0.4044
10	0.9360	0.0384	-0.4510	0.4605
11	0.9558	0.0295	-0.5200	0.5265
12	0.9696	0.0295	-0.5951	0.5993
13	0.9795	0.0164	-0.6951	0.6973
14	0.9874	0.0109	-0.8033	0.8041
15	0.9939	0.0057	-0.9332	0.9332
16	1	0	-	-

Table 1. The initial points on the revenue curve that are used to define the counter example distribution.



(a) Probability density function.



(b) Revenue curve.

Fig. 4. Properties of the distribution $\mathcal{F}$ defined in (6.2).

had the largest error in the previous step. The value in every subinterval is computed via a simple Trapezoidal method and a theoretically proven upper bound on the error is produced. The final result is an estimate $\hat{\ell}$ with the following guarantees

$$\hat{\ell} = 0.03185, \left|\hat{\ell} - \bar{\ell}\right| \le 0.00031 \implies \bar{\ell} \le 0.03216.$$

Finally from Table 1 we conclude that for the particular distribution $\mathcal{F}$ that we designed it holds that $\mathrm{OPT}_{\mathcal{F}} = 0.0501$ and hence we have that

$$\alpha^\star \le \frac{0.03216}{0.0501} \le 0.642$$

□

In the next plots we see the density and the revenue curve of the distribution that we used to prove Theorem 6.1.

7 CONCLUSIONS AND FUTURE WORK

We introduce a method for bounding the approximation ratio of algorithms that apply on a small set of independent samples from the same distribution. In such settings a very challenging part of the approximation ratio analysis is the identification of the worst distribution for which the performance of the algorithm is minimized. Our method is to design a convex relaxation that optimizes over the space of all possible distributions. The solution to the resulting convex program provides an upper bound on the approximation ratio and the solution to the dual convex program provides a proof for this upper bound.

In this paper we instantiate our method in the problem of estimating the approximation ratio that is achieved by the Empirical Revenue Maximization with two samples, compared to the optimal revenue achieved by Myerson's reserved price auction. Our method significantly improves the best known result and also gives us a lower bound on the approximation ratio.

Our method suggests the following future work:

- Potentially our method can be used to get a tight approximation or there is a relaxation gap. As far as we know if we run our suggested program with better discretization in machines with more computational power we could get even closer to the optimal approximation factor.
- Our SDP relaxation can be easily extended to the case wherein we are given k samples for any constant k. Of course the size of the program will be $O(n^{k+1})$, where n is the discretization parameter. Hence for larger values of k more computational power is needed. It is a very interesting and promising direction if we can get significantly better approximation ratio for $k \geq 3$ using our relaxation.
- There are many more problems in the AGT literature where the goal is to bound the approximation ratio of a mechanism that has access to a very small number of i.i.d. samples, e.g. [12], [13]. In these settings, using similar ideas we can provide SDP relaxations and it would be interesting to see if we can get better upper and lower bounds compared to the best known.

ACKNOWLEDGMENTS

The authors were supported by NSF Awards IIS-1741137, CCF-1617730 and CCF-1901292, by a Simons Investigator Award, by the DOE PhILMs project (No. DE-AC05-76RL01830), by the DARPA award HR0011199002, and by a Google Ph.D. Fellowship. C.D. thanks the Algorithms group at Google Research New York for hosting him in the early stages of this project, and in particular Vahab Mirrokni, Renato Paes Leme, Balu Sivan for useful discussions.

REFERENCES

[1] Noga Alon, Moshe Babaioff, Yannai A Gonczarowski, Yishay Mansour, Shay Moran, and Amir Yehudayoff. 2017. Submultiplicative Glivenko-Cantelli and uniform convergence of revenues. In *Advances in Neural Information Processing Systems*. 1656–1665.

[2] Moshe Babaioff, Yannai A. Gonczarowski, Yishay Mansour, and Shay Moran. 2018. Are Two (Samples) Really Better Than One?. In *Proceedings of the 2018 ACM Conference on Economics and Computation (EC '18)*.

[3] Maria-Florina Balcan, Tuomas Sandholm, and Ellen Vitercik. 2018. A general theory of sample complexity for multi-item profit maximization. In *Proceedings of the 2018 ACM Conference on Economics and Computation*. ACM, 173–174.

[4] Maria-Florina F Balcan, Tuomas Sandholm, and Ellen Vitercik. 2016. Sample complexity of automated mechanism design. In *Advances in Neural Information Processing Systems*. 2083–2091.

[5] Yang Cai and Constantinos Daskalakis. 2011. Extreme-value theorems for optimal multidimensional pricing. In *Proceedings of the 52nd Annual Symposium on Foundations of Computer Science*.

[6] Yang Cai and Constantinos Daskalakis. 2017. Learning Multi-item Auctions with (or without) Samples. In *the 58th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*.

[7] Richard Cole and Tim Roughgarden. 2014. The sample complexity of revenue maximization. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*.

[8] Nikhil R Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. 2016. The sample complexity of auctions with side information. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*. ACM, 426–439.

[9] Peerapong Dhangwatnotai, Tim Roughgarden, and Qiqi Yan. 2010. Revenue maximization with a single sample. In *Proceedings of the 11th ACM Conference on Electronic Commerce (EC)*.

[10] Shaddin Dughmi, Li Han, and Noam Nisan. 2014. Sampling and representation complexity of revenue maximization. In *Proceedings of the 10th International Conference on Web and Internet Economics (WINE)*.

[11] Edith Elkind. 2007. Designing and learning optimal finite support auctions. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics, 736–745.

[12] Hu Fu, Nicole Immorlica, Brendan Lucier, and Philipp Strack. 2015. Randomization beats second price as a prior-independent auction. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*.

[13] Kira Goldner and Anna R Karlin. 2016. A prior-independent revenue-maximizing auction for multiple additive bidders. In *International Conference on Web and Internet Economics*. Springer, 160–173.

[14] Yannai A Gonczarowski and Noam Nisan. 2017. Efficient empirical revenue maximization in single-parameter auction environments. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*.

[15] Yannai A Gonczarowski and S Matthew Weinberg. 2018. The Sample Complexity of Up-to-ε Multi-Dimensional Revenue Maximization. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 416–426.

[16] Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. 2015. Making the Most of Your Samples. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation (EC)*.

[17] Jamie Morgenstern and Tim Roughgarden. 2016. Learning simple auctions. In *Proceedings of the 30th Annual Conference on Learning Theory (COLT)*.

[18] Jamie H Morgenstern and Tim Roughgarden. 2015. On the pseudo-dimension of nearly optimal auctions. In *Proceedings of the the 29th Annual Conference on Neural Information Processing Systems (NIPS)*.

[19] Roger B Myerson. 1981. Optimal auction design. *Mathematics of operations research* 6, 1 (1981), 58–73.

[20] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. 2016. Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications* 169, 3 (2016), 1042–1068.

[21] John Riley and Richard Zeckhauser. 1983. Optimal selling strategies: When to haggle, when to hold firm. *The Quarterly Journal of Economics* 98, 2 (1983), 267–289.

[22] Tim Roughgarden and Okke Schrijvers. 2016. Ironing in the dark. In *Proceedings of the 2016 ACM Conference on Economics and Computation*. ACM, 1–18.