Extreme-Value Theorems for Optimal Multidimensional Pricing

Yang Cai[*]
EECS, MIT

Constantinos Daskalakis[†]
EECS, MIT

Abstract— We provide a Polynomial Time Approximation Scheme for the *multi-dimensional unit-demand pricing problem*, when the buyer's values are independent (but not necessarily identically distributed.) For all $\epsilon > 0$, we obtain a $(1+\epsilon)$-factor approximation to the optimal revenue in time polynomial, when the values are sampled from Monotone Hazard Rate (MHR) distributions, quasi-polynomial, when sampled from regular distributions, and polynomial in $n^{\text{poly}(\log r)}$, when sampled from general distributions supported on a set $[u_{min}, ru_{min}]$. We also provide an additive PTAS for all bounded distributions.

Our algorithms are based on novel extreme value theorems for MHR and regular distributions, and apply probabilistic techniques to understand the statistical properties of revenue distributions, as well as to reduce the size of the search space of the algorithm. As a byproduct of our techniques, we establish structural properties of optimal solutions. We show that, for all $\epsilon > 0$, $g(1/\epsilon)$ distinct prices suffice to obtain a $(1+\epsilon)$-factor approximation to the optimal revenue for MHR distributions, where $g(1/\epsilon)$ is a quasi-linear function of $1/\epsilon$ that does not depend on the number of items. Similarly, for all $\epsilon > 0$ and $n > 0$, $g(1/\epsilon \cdot \log n)$ distinct prices suffice for regular distributions, where n is the number of items and $g(\cdot)$ is a polynomial function. Finally, in the i.i.d. MHR case, we show that, as long as the number of items is a sufficiently large function of $1/\epsilon$, a single price suffices to achieve a $(1+\epsilon)$-factor approximation.

Our results represent significant progress to the single-bidder case of the multidimensional optimal mechanism design problem, following Myerson's celebrated work on optimal mechanism design [Myerson 1981].

1. INTRODUCTION

Here is a natural pricing problem: A seller has n items to sell to a buyer who is interested in buying a single item. The seller wants to maximize her profit from the sale, and wants to leverage stochastic knowledge she has about the buyer to achieve this goal. In particular, we assume that the seller has access to a distribution $\mathcal{F}$ from which the values $(v_1, \ldots, v_n)$ of the buyer for the items are drawn. Given this information, she needs to compute prices $p_1, \ldots, p_n$ for the items to maximize her revenue, assuming that the buyer is *quasi-linear*—i.e. will buy the item i maximizing $v_i - p_i$, as long as this difference is positive. Hence, the seller's expected payoff from a price vector $P = (p_1, \ldots, p_n)$ is

$$\mathcal{R}_P = \sum_{i=1}^{n} p_i \cdot \Pr\left[(i = \arg\max\{v_j - p_j\}) \wedge (v_i - p_i \geq 0)\right], \tag{1}$$

where we assume that the $\arg\max$ breaks ties in a consistent way, if there are multiple maximizers. A more sophisticated seller could try to improve her payoff by pricing lotteries over items, i.e. price randomized allocations (see [3],) albeit this may be less natural than item pricing.

While the problem looks simple, it exhibits a rich behavior depending on the nature of $\mathcal{F}$. For example, if $\mathcal{F}$ assigns the same value to all the items with probability 1, i.e. when the buyer always values all items equally, the problem degenerates to—what Economists call—a *single-dimensional* setting. In this setting, it is obvious that lotteries do not improve the revenue and that an optimal price vector should assign the same price to all items. This observation is a special case of a more general, celebrated result of Myerson [15] on optimal mechanism design (i.e. the multi-buyer version of the above problem, and generalizations thereof.) Myerson's result provides a closed-form solution to this generalized problem in a single sweep that covers many settings, but only works under the same limiting assumption that every buyer is single-dimensional, i.e. receives the same value from all the items (in general, the same value from all outcomes where she is provided service.)

Following Myerson's work, a large body of research in both Economics and Engineering has been devoted to extending this result to the *multi-dimensional setting*, i.e. when the buyers' values come from general distributions. And while there has been sporadic progress (see survey [14] and its references,) it appears that we are far from an optimal multi-dimensional mechanism, generalizing Myerson's result. In particular, there is no optimal solution known to even the single-buyer problem presented above. Even the ostensibly easier version of that problem, where the values of the buyer for the n items are independent and supported on a set of cardinality 2 appears challenging. Incidentally, the problem is trickier than what it originally seems, as several intuitive properties that one would expect from the optimal solution fail to hold. See the full version of this paper [5] for a discussion of these intricacies.

[*]Supported by NSF Awards CCF-0953960 and CCF-1101491.
[†]Supported by a Sloan Foundation Fellowship and NSF Awards CCF-0953960 (CAREER) and CCF-1101491.

Motivated by the importance of the problem to Economics, and intrigued by its simplicity and apparent difficulty, we devote this paper to the multi-dimensional pricing problem. Our main contribution is to develop the *first near-optimal algorithms for this problem,* when the buyer's values are independent (but not necessarily identically distributed) random variables.

Previous work on this problem by Chawla et al. [6], [7] provides a factor 2 approximation to the revenue achieved by the optimal price vector. The elegant observation enabling this result is to consider the following mental experiment: suppose that the unit-demand buyer is split into n "copies" $t_1, \ldots, t_n$. Copy t_i is only interested in item i and her value for that item is drawn from the distribution $\mathcal{F}_i$ (where $\mathcal{F}_i$ is the marginal of $\mathcal{F}$ on item i), independently from the values of the other copies. On the other hand, the seller has the same feasibility constraints as before: only one item can be sold in this auction. It is intuitively obvious and can be formally established that the seller in the latter scenario is in better shape: there is more competition in the market and this can be exploited to extract more revenue. So the revenue of the seller in the original scenario can be upper bounded by the revenue in the hypothetical scenario. Moreover, the latter is a single-parameter setting; hence, we understand exactly how its optimal revenue behaves by Myerson's result. So we can go back to our original setting and design a mechanism whose revenue comes close to Myerson's revenue in the hypothetical scenario. Using this approach, [7] obtains a 2-approximation to the optimal revenue. Moreover, if the distributions $\{\mathcal{F}_i\}_i$ are *regular* (this is a commonly studied class of distributions in Economics,) the corresponding price vector can be computed efficiently.

Nevertheless, there is an inherent loss in the approach outlined above, as the revenue obtained by the sought-after mechanism will eventually be compared to a revenue that is not the optimal achievable revenue in the real setting, but the optimal revenue in a hypothetical setting; and as far as we know this could be up to a factor of 2 larger than the real one. So it could be that this approach is inherently limited to constant factor approximations. We are interested instead in computationally efficient pricing mechanisms that achieve a $(1-\epsilon)$-fraction of the optimal revenue, for arbitrarily small ϵ. We show

Theorem 1 (PTAS for MHR Distributions)**.** *For all $\epsilon > 0$, there is a Polynomial Time Approximation Scheme [1] for computing a price vector whose revenue is a $(1+\epsilon)$-factor approximation to the optimal revenue, when the values of the buyer are independent and drawn from Monotone Hazard Rate distributions. (This is a commonly studied class of distributions in Economics—see Section 2.) For all $\epsilon > 0$, the algorithm runs in time $n^{\text{poly}(1/\epsilon)}$.*

Theorem 2 (Quasi-PTAS for Regular Distributions)**.** *For all $\epsilon > 0$, there is a Quasi-Polynomial Time Approximation Scheme [2] for computing a price vector whose revenue is a $(1+\epsilon)$-factor approximation to the optimal revenue, when the values of the buyer are independent and drawn from regular distributions. (These contain MHR and are also commonly studied in Economics—see Section 2.) For all $\epsilon > 0$, the algorithm runs in time $n^{\text{poly}(\log n, 1/\epsilon)}$.*

Theorem 3 (General Algorithm)**.** *For all $\epsilon > 0$, there is an algorithm for computing a price vector whose revenue is a $(1+\epsilon)$-factor approximation to the optimal revenue, whose running time is $n^{\text{poly}(\frac{1}{\epsilon}, \log r)}$ when the values of the buyer are independent and distributed in an interval $[u_{min}, ru_{min}]$.* [3]

Theorem 4 (Additive PTAS–General Distributions)**.** *For all $\epsilon > 0$, there is a PTAS for computing a price vector whose revenue is within an additive ϵ of the optimal revenue, when the values of the buyer are independent and distributed in $[0, 1]$.*

Structural Theorems: Our approach is different than that of [6], [7] in that we study directly the optimal revenue (as a random variable,) rather than only relating its expectation to a benchmark that may be off by a constant factor. Clearly, the optimal revenue is a function of the values (which are random) and the optimal price vector (which is unknown). Hence it may be hard to pin down its distribution exactly. Nevertheless, we manage to understand its statistical properties sufficiently to deduce the following interesting structural theorems.

Theorem 5 (**Structural 1:** A Constant Number of Distinct Prices Suffice for MHR Distributions)**.** *There exists a (quasi-linear) function $g(\cdot)$ such that, for all $\epsilon, n > 0$, $g(1/\epsilon)$ distinct prices suffice for a $(1+\epsilon)$-approximation to the optimal revenue when the buyer's values for the n items are independent and MHR. These distinct prices can be computed efficiently from the value distributions.*

Theorem 6 (**Structural 2:** A Polylog Number of Distinct Prices Suffice for Regular Distributions)**.** *There exists a (polynomial) function $g(\cdot)$ such that, for all $\epsilon, n > 0$, $g(1/\epsilon \cdot \log n)$ distinct prices suffice for a $(1+\epsilon)$-approximation to the optimal revenue, when the buyer's values for the n items are independent and regular. These prices can be computed efficiently from the value distributions.*

[1] Recall that a Polynomial Time Approximation Scheme (PTAS) is a family of algorithms $\{\mathcal{A}_\epsilon\}_\epsilon$, indexed by a parameter $\epsilon > 0$, such that for every fixed $\epsilon > 0$, $\mathcal{A}_\epsilon$ runs in time polynomial in the size of its input.

[2] Recall that a Quasi Polynomial Time Approximation Scheme (Quasi-PTAS) is a family of algorithms $\{\mathcal{A}_\epsilon\}_\epsilon$, indexed by a parameter $\epsilon > 0$, such that for every fixed $\epsilon > 0$, $\mathcal{A}_\epsilon$ runs in time quasi-polynomial in the size of its input.

[3] We point out that a straightforward application of the discretization proposed by Nisan (see [6]) or Hartline and Koltun [12] would only give a $\left(\frac{1}{\epsilon}\log r\right)^{O(n)}$-time algorithm.

Theorem 5 shows that, when the values are MHR independent, then only the desired approximation ϵ dictates the number of distinct prices that are necessary to achieve a $(1+\epsilon)$-approximation to the optimal revenue, and the number of items n as well as the range of the distributions *are irrelevant* (!) Theorem 6 generalizes this to a mild dependence on n for regular distributions. Establishing these theorems is quite challenging, as it relies on a deep understanding of the properties of the tails of MHR and regular distributions. For this purpose, we develop novel *extreme value theorems* for these classes of distributions (Theorems 12 and 14 in Sections 3 and 4 respectively.) Our theorems bound the size of the tail of the *maximum* of n independent (but not necessarily identically distributed) random variables, which are MHR or regular respectively, and are instrumental in establishing the following truncation property: truncating all the values into a common interval of the form $[\alpha, \mathrm{poly}(1/\epsilon)\alpha]$ in the MHR case, and $[\alpha, \mathrm{poly}(n, 1/\epsilon)\alpha]$ in the regular case, for some α that depends on the value distributions, only loses a fraction of ϵ of the optimal revenue. This is quite remarkable, especially in the case where the value distributions are non-identical. Why should most of the contribution to the optimal revenue come from a restricted set as above, when each of the underlying value distributions may concentrate on different supports? We expect that our extreme value theorems will be useful in future work, and indeed they have already been used [9]. As a final remark, we would like to point out that extreme value theorems *have* been obtained in Statistics for large classes of distributions [10], and indeed those theorems have been applied earlier in optimal mechanism design [2]. Nevertheless, known extreme value theorems are typically asymptotic, only hold for maxima of i.i.d. random variables, and are not known to hold for all MHR or regular distributions.

Covers of Revenue Distributions: Our structural theorems enable us to significantly reduce the search space for an (approximately) optimal price vector. Indeed, if the values are i.i.d., we can easily establish Theorems 1, 2 and 3, using the aforementioned structural theorems and exploiting the symmetry across items. Nevertheless, our value distributions are not necessarily identically distributed, so the search space remains exponentially large even for the MHR case, where a constant (function of ϵ only) number of distinct prices suffice by Theorem 5. Even if there are only 2 possible prices per item, but the items are not identically distributed, how can we decide efficiently what price to assign to each item?

The natural approach would be to cluster the distributions into a small number of buckets, containing distributions with similar statistical properties, and proceed to treat all items in a bucket as essentially identical. However, the problem at hand is not sufficiently smooth for us to perform such bucketing into a small number of buckets, and several intuitive bucketing approaches fail. We can obtain a delicate discretization of the support of the distributions into a small set, but cannot discretize the probabilities used by these distributions into coarse-enough accuracy, arriving at an impasse with discretization ideas.

Our next conceptual idea is to shift the focus of attention from the space of *input value distributions*, which is inherently exponential, to the space of all possible *revenue distributions*, which are *scalar* random variables. (As we mentioned earlier, the revenue from a given price vector can be viewed as a random variable that depends on the values.) There are still exponentially many possible revenue distributions (one for each price vector,) but we find a way to construct a sparse δ-cover of this space under the total variation distance between distributions. The cover is implicit, i.e. it has no succinct closed-form description. We argue instead that it can be produced by a dynamic program, which considers prefixes of the items and constructs sub-covers for the revenue distributions induced by these prefixes, pruning down the size of the cover before growing it again to include the next item. Once a cover of the revenue distributions is obtained in this way, we argue that there is only a δ-fraction of revenue lost by replacing the optimal revenue distribution with its proxy in the cover. The high-level structure of the argument is provided in Section 6, and the details are in Section 7. The proofs of our algorithmic results (Theorems 1, 2, 3 and 4) can be found in the full paper.

Extensions: A natural conjecture is that, when the distributions are not widely different, a single price should suffice for extracting a $(1-\epsilon)$-fraction of the optimal revenue; that is, as long as there is a sufficient number of items for sale. We show such a result in the case that the buyer's values are i.i.d. according to a MHR distribution.

Theorem 7 (**Structural 3 (i.i.d.):** A Single Price Suffices for MHR Distributions)**.** *There is a function $g(\cdot)$ such that, for any $\epsilon > 0$, if the number of items $n > g(1/\epsilon)$ then a single price suffices for a $(1+\epsilon)$-factor approximation to the optimal revenue, if the buyer's values are i.i.d. and MHR.*

Another interesting byproduct of our techniques is that any constant-factor approximation to the optimal pricing can be converted into a PTAS or a quasi-PTAS respectively in the case of MHR or regular value distributions. This result is a direct product of our extreme value theorems, which can be boot-strapped with a constant factor approximation to OPT. Having such approximation would obviate the need to use our generic algorithm, outlined in the proofs of Theorems 5 and 6.

Theorem 8 (Constant Factor to Near-Optimal Approximation)**.** *If we have a constant-factor approximation to the optimal revenue of an instance of the pricing problem* *where the values are either MHR or regular, we can use this to speed-up our algorithms of Theorems 1 and 2.*

Future and Related Work. In conclusion, this paper provides the first near-optimal efficient algorithms for interesting instances of the multi-dimensional mechanism design problem, for a unit-demand bidder whose values are independent (but not necessarily identically distributed.) Our results provide algorithmic, structural and probabilistic insights into the properties of the optimal deterministic mechanism for the case of MHR, regular, and more general distributions. It would be interesting to extend our results (algorithmic and/or structural) to more general distributions, to mechanisms that price lotteries over items [18], [3], to bundle-pricing [13] and to budgets [1], [17]. We can certainly obtain such extensions, albeit when sizes of lotteries, bundles, etc. are a constant. We believe that our extreme value theorems, and our probabilistic view of the problem in terms of revenue distributions will be helpful in obtaining more general results. We also leave the complexity of the exact problem as an open question, and conjecture that it is NP-hard, referring the reader to [4] for hardness results in the case of correlated distributions.

Finally, it is important to solve the multi-bidder problem, extending Myerson's celebrated mechanism to the multi-dimensional setting, and the results of [1], [7] beyond constant factor approximations. In recent work, Daskalakis and Weinberg [9] have made progress in this front obtaining efficient mechanisms for multi-bidder multi-item auctions. These results are neither subsumed, nor subsume the results in the present paper. Indeed, we are more general here in that we allow the buyer to have values for the items that are not necessarily i.i.d., an assumption needed in [9] if the number of items is large. On the other hand, we are less general in that (a) we solve the single-bidder problem and (b) are near-optimal with respect to all deterministic (i.e. item-pricing), but not necessarily randomized (lottery-pricing) mechanisms. Strikingly, the techniques of the present paper are essentially orthogonal to those of [9]. The approach of [9] uses randomness to symmetrize the solution space, coupling this symmetrization with Linear Programming formulations of the problem. Our paper takes instead a probabilistic approach, developing extreme value theorems to characterize the optimal solution, and designing covers of revenue distributions to obtain efficient algorithmic solutions. It is tempting to conjecture that our approach here, combined with that of [9] would lead to more general results. Indeed, our extreme value theorems found use in [9], but we expect that significant technical work is required to go forward.

2. Preliminaries

For a random variable X we denote by $F_X(x)$ the cumulative distribution function of X, and by $f_X(x)$ its probability density function. We also let $u_{min}^X = \sup\{x | F_X(x) = 0\}$ and $u_{max}^X = \inf\{x | F_X(x) = 1\}$. u_{max}^X may be $+\infty$, but we assume that $u_{min}^X \geq 0$, since the distributions we consider in this paper represent value distributions of items. Moreover, we often drop the subscript or superscript of X, if X is clear from context. A natural question is how distributions are provided as input to an algorithm (explicitly or with oracle access). We discuss this technical issue in the appendix. We also define precisely what it means for an algorithm to be "efficient" in each case. We continue with the precise definition of *Monotone Hazard Rate (MHR)* and *Regular* distributions, which are both commonly studied classes of distributions in Economics.

Definition 9 (Monotone Hazard Rate Distribution)**.** *We say that a one-dimensional differentiable distribution F has* Monotone Hazard Rate*, shortly* MHR*, if $\frac{f(x)}{1-F(x)}$ is non-decreasing in $[u_{min}, u_{max}]$.*

Definition 10 (Regular Distribution)**.** *A one-dimensional differentiable distribution F is called* regular *if $x - \frac{1-F(x)}{f(x)}$ is non-decreasing in $[u_{min}, u_{max}]$.*

It is worth noticing that all MHR distributions are also regular distributions, but there are regular distributions that are not MHR. The family of MHR distributions includes such familiar distributions as the Normal, Exponential, and Uniform distributions. The family of regular distributions contains a broader range of distributions, such as fat-tail distributions $f_X(x) \sim x^{-(1+\alpha)}$ for $\alpha \geq 1$ (which are not MHR). In the full version of the paper, we establish important properties of MHR and regular distributions. These properties are instrumental in establishing our extreme value theorems (Theorems 12 and 14 in the following sections).

We conclude the section by defining two computational problems that we use in the next sections. For the types of value distributions we consider, we can show that these problems are well-defined (i.e. have finite optimal solutions.)

PRICE: **Input:** A collection of mutually independent random variables $\{v_i\}_{i=1}^n$, and some $\epsilon > 0$.
Output: A vector of prices $(p_1, \ldots, p_n)$ such that the expected revenue $\mathcal{R}_P$ under this price vector, defined as in Eq. (1), is within a $(1+\epsilon)$-factor of the optimal revenue achieved by any price vector.

RESTRICTEDPRICE: **Input:** A collection of mutually independent random variables $\{v_i\}_{i=1}^n$, a discrete set $\mathcal{P} \subset \mathbb{R}_+$, and some $\epsilon > 0$.
Output: A vector of prices $(p_1, \ldots, p_n) \in \mathcal{P}^n$ such that the expected revenue $\mathcal{R}_P$ under this price vector is within a $(1 + \epsilon)$-factor of the optimal revenue achieved by any vector in $\mathcal{P}^n$.

3. Extreme Values of MHR Distributions

We reduce the problem of finding a near-optimal price vector for MHR distributions to finding a near-optimal price vector for value distributions supported on a common, balanced interval, where the imbalance of the interval is only a function of the desired approximation $\epsilon > 0$. More precisely,

Theorem 11 (From MHR to Balanced Distributions)**.** *Let $\mathcal{V} = \{v_i\}_{i\in[n]}$ be a collection of mutually independent (but not necessarily identically distributed) MHR random variables. Then there exists some $\beta = \beta(\mathcal{V}) > 0$ such that for all $\epsilon \in (0, 1/4)$, there is a reduction from $\text{PRICE}(\mathcal{V}, c\epsilon\log(\frac{1}{\epsilon}))$ to $\text{PRICE}(\tilde{\mathcal{V}}, \epsilon)$, where $\tilde{\mathcal{V}} := \{\tilde{v}_i\}_i$ is a collection of mutually independent random variables supported on the set $[\frac{\epsilon}{2}\beta, 2\log\frac{1}{\epsilon}\beta]$, and c is some absolute constant.*

Moreover, β is efficiently computable from the distributions of the X_i's (whether we are given the distributions explicitly, or we have oracle access to them,) and for every ϵ the running time of the reduction is polynomial in the size of the input and $\frac{1}{\epsilon}$. In particular, if we have oracle access to the distributions of the v_i's, then the forward reduction produces oracles for the distributions of the $\tilde{v}_i$'s, which run in time polynomial in n, $1/\epsilon$, the input to the oracle and the desired oracle precision.

We discuss the essential elements of this reduction below. Most crucially, the reduction is enabled by the following characterization of the extreme values of a collection of independent, but not necessarily identically distributed, MHR distributions.

Theorem 12 (Extreme Values of MHR distributions)**.** *Let $X_1, \ldots, X_n$ be a collection of independent (but not necessarily identically distributed) random variables whose distributions are MHR. Then there exists some anchoring point β such that $\Pr[\max_i\{X_i\} \geq \beta/2] \geq 1 - \frac{1}{\sqrt{e}}$ and for all $\epsilon \in (0, 1/4)$,*

$$\int_{2\beta\log 1/\epsilon}^{+\infty} t \cdot f_{\max_i\{X_i\}}(t)dt \leq 36\beta\epsilon\log 1/\epsilon. \quad (2)$$

Moreover, β is efficiently computable from the distributions of the X_i's (whether we are given the distributions explicitly, or we have oracle access to them.)

Theorem 12 shows that, for all ϵ, at least a $(1 - O(\epsilon\log\frac{1}{\epsilon}))$-fraction of $\mathbb{E}[\max_i X_i]$ is contributed to by values that are no larger than $\mathbb{E}[\max_i X_i] \cdot \log\frac{1}{\epsilon}$. Our result is quite surprising, especially for the case of non-identically distributed MHR random variables. Why should most of the contribution to $\mathbb{E}[\max_i X_i]$ come from values that are close (*within a function of ϵ only*) to the expectation, when the underlying random variables X_i may concentrate on widely different supports? To obtain the theorem one needs to understand how the tails of the distributions of a collection of independent but not necessarily identically distributed MHR random variables contribute to the expectation of their maximum. Our proof technique is rather intricate, defining a tournament between the tails of the distributions. Each round of the tournament ranks the distributions according to the size of their tails, and eliminates the lightest half. The threshold β is then obtained by some side-information that the algorithm records in every round.

Here we only describe the tournament algorithm for computing β, postponing the proof of Theorem 12 to the full paper. We start with some useful notation. For all $i = 1, \ldots, n$, we denote by F_i the distribution of variable X_i. We also let $\alpha_m^{(i)} := \inf\{x | F_i(x) \geq 1 - \frac{1}{m}\}$, for all $m \geq 1$. Moreover, we assume that n is a power of 2. If not, we can always include at most n additional random variables that are detreministically 0, making the total number of variables a power of 2.

We proceed with the algorithm. At a high level, the algorithm proceeds in $O(\log n)$ rounds, indexed by $t \in \{0, \ldots, \log n\}$, eliminating half of the variables at each round. The way the elimination works is as follows. In round t, we compute for each of the variables that have survived so far the threshold $\alpha_{n/2^t}$ beyond which the size of the tail of their distribution becomes smaller than $1/(n/2^t)$. We then sort these thresholds and eliminate the bottom half of the variables, recording the threshold of the last variable that survived this round. The maximum of these records among the $\log n$ rounds of the algorithm is our β. The pseudocode of the algorithm is given below. Given that we may only be given oracle access to the distributions $\{F_i\}_{i\in[n]}$, we allow some slack $\eta \leq \frac{1}{2}$ in the computation of our thresholds so that the computation is efficient. If we know the distributions explicitly, the description of the algorithm simplifies to the case $\eta = 0$.

Algorithm 1 Algorithm for finding β

Define the permutation of the variables $\pi_0(i) = i$, $\forall i \in [n]$, and the set of remaining variables $Q_0 = [n]$.
for $t := 0$ to $\log n - 1$ **do**
    For all $j \in [n/2^t]$, compute some $x_{n/2^t}^{(\pi_t(j))} \in [1 - \eta, 1 + \eta] \cdot \alpha_{n/2^t}^{(\pi_t(j))}$, for a small constant $\eta \in [0, 1/2)$
    Sort these $n/2^t$ numbers in decreasing order π_{t+1} such that $x_{n/2^t}^{(\pi_{t+1}(1))} \geq x_{n/2^t}^{(\pi_{t+1}(2))} \geq \ldots \geq x_{n/2^t}^{(\pi_{t+1}(n/2^t))}$
    $Q_{t+1} := \{\ \pi_{t+1}(i) \mid i \leq n/2^{t+1}\ \}$
    $\beta_t := x_{n/2^t}^{(\pi_{t+1}(n/2^{t+1}))}$
end for
Compute $x_2^{(\pi_{\log n}(1))} \in [1 - \eta, 1 + \eta] \cdot \alpha_2^{(\pi_{\log n}(1))}$
Set $\beta_{\log n} := x_2^{(\pi_{\log n}(1))}$
Output $\beta := \max_t \beta_t$

Given our understanding of the extreme values of MHR distributions, our reduction of Theorem 11 from MHR to

Balanced distributions proceeds in the following steps:

- We start with the computation of the threshold β specified by Theorem 12. This computation can be done efficiently using the tournament algorithm described above. Given that $\Pr[\max_i\{X_i\} \geq \beta/2]$ is bounded away from 0, β provides a lower bound to the optimal revenue. (See the full paper for the precise lower bound we obtain.) Such lower bound is useful as it implies that, if our transformation loses revenue that is a small fraction of β, this corresponds to a small fraction of optimal revenue lost.
- Next, using (2) we show that, for all $\epsilon > 0$, if we restrict the prices to lie in the balanced interval $[\epsilon \cdot \beta, 2\log(\frac{1}{\epsilon}) \cdot \beta]$, we only lose a $O(\epsilon \log 1/\epsilon)$ fraction of the optimal revenue;
- Finally, we show that we can efficiently transform the given MHR random variables $\{v_i\}_{i\in[n]}$ into a new collection of random variables $\{\tilde{v}_i\}_{i\in[n]}$ that take values in $[\frac{\epsilon}{2} \cdot \beta, 2\log(\frac{1}{\epsilon}) \cdot \beta]$ and satisfy the following: a near-optimal price vector for the setting where the buyer's values are distributed as $\{\tilde{v}_i\}_{i\in[n]}$ can be efficiently transformed into a near-optimal price vector for the original setting, i.e. where the buyer's values are distributed as $\{v_i\}_{i\in[n]}$.

4. Extreme Values of Regular Distributions

Our goal is to reduce the problem of finding a near-optimal pricing for a collection of independent (but not necessarily identical) *regular* value distributions to the problem of finding a near-optimal pricing for a collection of independent distributions, which are supported on a common finite interval $[u_{min}, u_{max}]$ with $u_{max}/u_{min} \leq 16n^8/\epsilon^4$, where n is the number of distributions and ϵ is the desired approximtion. It is important to notice that our bound on the ratio u_{max}/u_{min} does not depend on the distributions at hand, just their number and the required approximation. We also emphasize that the input regular distributions may be supported on $[0, +\infty)$, so it is a priori not clear if we can truncate these distributions to any finite set (even of exponential imbalance) without losing revenue.

Theorem 13 (Reduction from Regular to $Poly(n)$-Balanced Distributions)**.** *Let $\mathcal{V} = \{v_i\}_{i\in[n]}$ be a collection of mutually independent (but not necessarily identically distributed)* regular *random variables. Then there exists some $\alpha = \alpha(\mathcal{V}) > 0$ such that, for any $\epsilon \in (0, 1)$, there is a reduction from* $\textsc{Price}(\mathcal{V}, \epsilon)$ *to* $\textsc{Price}(\tilde{\mathcal{V}}, \epsilon - \Theta(\epsilon/n))$*, where $\tilde{\mathcal{V}} = \{\tilde{v}_i\}_{i\in[n]}$ is a collection of mutually independent random variables that are supported on* $[\frac{\epsilon\alpha}{4n^4}, \frac{4n^4\alpha}{\epsilon^3}]$.

Moreover, we can compute α in time polynomial in n and the size of the input (whether we have the distributions of the v_i's explicitly, or have oracle access to them.) For all ϵ, the reduction runs in time polynomial in n, $1/\epsilon$ and the size of the input. In particular, if we have oracle access to the distributions of the v_i's, then the forward reduction produces oracles for the distributions of the $\tilde{v}_i$'s, which run in time polynomial in n, $1/\epsilon$, the input to the oracle and the desired oracle precision.

Our reduction is based on the following extreme value theorem for regular distributions.

Theorem 14 (Homogenization of the Extreme Values of Regular Distributions)**.** *Let $\{X_i\}_{i\in[n]}$ be a collection of mutually independent (but not necessarily identically distributed) regular random variables, where $n \geq 2$. Then there exists some $\alpha = \alpha(\{X_i\}_i)$ such that:*

1) *α has the following "anchoring" properties:*
   - *for all $\ell \geq 1$, $\Pr[X_i \geq \ell\alpha] \leq 2/(\ell n^3)$, for all $i \in [n]$;*
   - *$\alpha/n^3 \leq c \cdot \max_z(z \cdot \Pr[\max_i\{X_i\} \geq z])$, where c is an absolute constant.*
2) *for all $\epsilon \in (0, 1)$, the tails beyond $\frac{2n^2\alpha}{\epsilon^2}$ can be "homogenized", i.e.*
   - *for any integer $m \leq n$, thresholds $t_1, \ldots, t_m \geq t \geq \frac{2n^2\alpha}{\epsilon^2}$, and index set $\{a_1, \ldots, a_m\} \subseteq [n]$:*

$$\begin{aligned}&\sum_{i=1}^{m} t_i \Pr[X_{a_i} \geq t_i] \\ &\leq \left(t - \frac{2\alpha}{\epsilon}\right) \cdot \Pr\left[\max_i\{X_{a_i}\} \geq t\right] \\ &\quad + \frac{7\epsilon}{n} \cdot \left(\frac{2\alpha}{\epsilon} \cdot \Pr\left[\max_i\{X_{a_i}\} \geq \frac{2\alpha}{\epsilon}\right]\right).\end{aligned}$$

Finally, α is efficiently computable from the distributions of the X_i's (whether we are given the distributions explicitly, or have oracle access to them.)

We discuss the meaning of our homogenization theorem in Appendix B. Given this theorem, Theorem 13 is obtained as follows.

- First, we compute the threshold α specified in Theorem 14. This can be done efficiently as stated in Theorem 14. Now given the second anchoring property of α, we obtain an $\Omega(\alpha/n^3)$ lower bound to the optimal revenue. Such a lower bound is useful as it implies that we can ignore prices below some $O(\epsilon\alpha/n^3)$.
- Next, using our homogenization Theorem 14, we show that if we restrict a price vector to lie in $[\epsilon\alpha/n^4, 2n^2\alpha/\epsilon^2]^n$, we only lose a $O(\frac{\epsilon}{n})$ fraction of the optimal revenue.
- Finally, we show that we can efficiently transform the input regular random variables $\{v_i\}_{i\in[n]}$ into a new collection of random variables $\{\tilde{v}_i\}_{i\in[n]}$ that are supported on $[\frac{\epsilon\alpha}{4n^4}, \frac{4n^4\alpha}{\epsilon^3}]$ and satisfy the following: a near-optimal price vector for when the buyer's values are distributed as $\{\tilde{v}_i\}_{i\in[n]}$ can be efficiently transformed into a near-optimal price vector for when the buyer's values are distributed as $\{v_i\}_{i\in[n]}$.

5. From Continuous to Discrete Distributions

The expected revenue can be sensitive even to small perturbations of the prices and the probability distributions. So it is a priori not clear whether there is a coarse discretization of the input distributions and the search space for price vectors that does not cost us a lot of revenue. We show that there is such discretization, but needs to be done carefully. Our discretization result is summarized in Theorem 15. Note that the discretization we obtain does not eliminate the exponentiality of the search space or the space of input distributions.

Theorem 15 (Price/Value Distribution Discretization)**.** *Let $\mathcal{V} = \{v_i\}_{i\in[n]}$ be a collection of mutually independent random variables supported on a finite set $[u_{min}, u_{max}] \subset \mathbb{R}_+$, and let $r = \frac{u_{max}}{u_{min}} \geq 1$. For any $\epsilon \in \left(0, \frac{1}{(4\lceil\log r\rceil)^{1/6}}\right)$, there is a reduction from* Price$(\mathcal{V}, \epsilon)$ *to* RestrictedPrice$(\hat{\mathcal{V}}, \mathcal{P}, \Theta(\epsilon^8))$*, where*

- *$\hat{\mathcal{V}} = \{\hat{v}_i\}_{i\in[n]}$ is a collection of mutually independent random variables that are supported on a common set of cardinality $O\left(\frac{\log r}{\epsilon^{16}}\right)$;*
- *$|\mathcal{P}| = O\left(\frac{\log r}{\epsilon^{16}}\right)$.*

Moreover, assuming that the set $[u_{min}, u_{max}]$ is specified in the input, [4] *we can compute the (common) support of the distributions of the variables $\{\hat{v}_i\}_i$ as well as the set of prices $\mathcal{P}$ in time polynomial in $\log u_{min}$, $\log u_{max}$ and $1/\epsilon$. We can also compute the distributions of the variables $\{\hat{v}_i\}_{i\in[n]}$ in time polynomial in the size of the input and $1/\epsilon$, if we have the distributions of the variables $\{v_i\}_{i\in[n]}$ explicitly. If we have oracle access to the distributions of the variables $\{v_i\}_{i\in[n]}$, we can construct an oracle for the distributions of the variables $\{\hat{v}_i\}_{i\in[n]}$, running in time polynomial in $\log u_{min}$, $\log u_{\max}$, $1/\epsilon$, the input to the oracle and the desired precision.*

That prices can be discretized follows immediately from a discretization lemma attributed to Nisan [6] (see also a related discretization in [12].) The discretization of the value distributions is inspired by Nisan's lemma, but requires an intricate twist in order to reduce the size of the support to be linear in $\log r$ rather than linear in $r^2 \log r$ which is what a straightforward modification of the lemma gives. (Indeed, quite some effort is needed to get the former bound.) For more details, we refer interested readers to our full paper.

6. Probabilistic Covers of Revenue Distributions

Let $\mathcal{V} := \{v_i\}_i$ be an instance of Price, where the v_i's are mutually independent random variables distributed on a finite set $[u_{min}, u_{max}]$ according to distributions $\{F_i\}_i$, and let $\mathcal{R}_{OPT}$ be the optimal expected revenue for $\mathcal{V}$. Our goal is to compute a price vector with expected revenue $(1-\epsilon)\mathcal{R}_{OPT}$. Theorem 15 of Section 5 provides an efficient reduction of this problem to the $(1-\delta)$-approximation of a discretized problem, where both the values and the prices come from discrete sets whose cardinality is $O(\log r/\delta^2)$, where $r = \frac{u_{max}}{u_{min}}$ and $\delta = O(\epsilon^8)$. For convenience, we denote by $\{\hat{F}_i\}_i$ the resulting discretized distributions, by $\{\hat{v}_i\}_i$ a collection of mutually independent random variables distributed according to the $\hat{F}_i$'s, by $\{v^{(1)}, v^{(2)}, \ldots, v^{(k_1)}\}$ the (common) support of all the $\hat{F}_i$'s, and by $\{p^{(1)}, p^{(2)}, \ldots, p^{(k_2)}\}$ the set of available price levels, where both k_1 and k_2 are $O(\log r/\delta^2)$. It is worth noting that the set of prices obtained from Theorem 15 satisfies $\min\{p^{(i)}\} \geq u_{min}/(1+\delta)$ and $\max\{p^{(i)}\} \leq u_{max}$, and that these prices are points of a geometric sequence of ratio $1/(1-\delta^2)$.

Having discretized the support sets of values and prices, a natural idea that one would like to use to go forward would be to further discretize the distributions $\{\hat{F}_i\}_i$ by rounding the probabilities they assign to every point in their support to integer multiples of some fraction $\sigma = \sigma(\epsilon, r) > 0$, i.e. a fraction that does not depend on n. If such discretization were feasible, the problem would be greatly simplified. For example, if additionally r were an absolute constant or a function of ϵ only, there would only be a constant number of possible value distributions (as both the cardinality of the support of the distributions and the number of available probability levels would be a function of ϵ only.) In such case, we could try to develop an algorithm tailored to a constant number of available value distributions. This is still not easy to do (as we don't even know how to solve the i.i.d. case of our problem), but is definitely easier to dream of. Nevertheless, the approach breaks down as preserving the revenue while doing a coarse rounding of the probabilities appears difficult, and the best discretization we can obtain requires accuracy which is inverse polynomial in n.

Given the apparent impasse towards eliminating the exponentiality from the input space of our problem, our solution evolves in a radically different direction. To explain our approach, let us view our problem in the graphical representation of Figure 1. Circuit C takes as input a price vector $p_1, \ldots, p_n$ and outputs the distribution $F_{\hat{R}_P}$ of the revenue of the seller under this price vector. Indeed, the revenue of the seller is a random variable $\hat{R}_P$ whose value depends on the variables $\{\hat{v}_i\}_{i\in[n]}$. So in order to compute the distribution of the revenue the circuit also uses the distributions $\{\hat{F}_i\}_{i\in[n]}$, which are hard-wired into the circuit. Let us denote the expectation of $\hat{R}_P$ as $\hat{\mathcal{R}}_P$.

Given our restriction of the prices to the finite set $\{p^{(1)}, p^{(2)}, \ldots, p^{(k_2)}\}$, there are k_2^n possible inputs to the circuit, and a corresponding k_2^n number of possible revenue distributions that the circuit can produce. Our main conceptual idea is this: *instead of worrying about*

[4] The requirement that the set $[u_{min}, u_{max}]$ is specified as part of the input is only relevant if we have oracle access to the distributions of the v_i's, as if we have them explicitly we can easily find $[u_{min}, u_{max}]$.

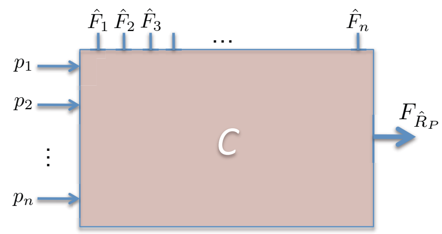

Figure 1. The Revenue Distribution.

the set of inputs to circuit C, we focus on the revenue distribution directly, constructing a probabilistic cover (under an appropriate metric) of all the possible revenue distributions that can be output by the circuit. The two crucial properties of our cover are the following: (a) it has cardinality $O(n^{\text{poly}(\frac{1}{\epsilon}, \log r)})$, and (b) for any possible revenue distribution that the circuit may output, there exists a revenue distribution in our cover with approximately the same expectation.

Details of the Cover.: At a high level, the way we construct our cover is via dynamic programming, whose steps are interleaved with coupling arguments pruning the size of the DP table before proceeding to the next step. Intuitively, our dynamic program sweeps the items from 1 through n, maintaining a cover of the revenue distributions produced by all possible pricings on a prefix of the items. More precisely, for each prefix of the items, our DP table keeps track of all possible feasible collections of $k_1 \times k_2$ probability values, where Pr_{i_1,i_2}, $i_1 \in [k_1], i_2 \in [k_2]$, denotes the probability that the item with the largest value-minus-price gap (i.e. the item of the prefix that would have been sold in a sale that only sales the prefix of items) has value $v^{(i_1)}$ for the buyer and is assigned price $p^{(i_2)}$ by the seller. I.e. we memoize all possible (winning-value, winning-price) distributions that can arise from each prefix of items. The reasons we decide to memoize these distributions are the following:

- First, if we have these distributions, we can compute the expected revenue that the seller would obtain, if we restricted our sale to the prefix of items.
- Second, when our dynamic program considers assigning a particular price to the next item, then having the (winning-value, winning-price) distribution on the prefix suffices to obtain the new (winning-value,winning-price) distribution that also includes the next item. I.e., if we know these distributions, we do not need to keep track of anything else in the history to keep going. Observe that it is crucial here to maintain the joint distribution of both the winning-value and the winning-price, rather than just the distribution of the winning-price.
- In the end of the program, we can look at all feasible (winning-value,winning-price) distributions for the full set of items to find the one achieving the best revenue; we can then follow back-pointers stored in our DP table to uncover a price vector consistent with the optimal distribution.

All this is both reasonable, and fun, but thus far we have achieved nothing in terms of reducing the number of distributions $F_{\hat{R}_p}$ in our cover. Indeed, there could be exponentially many (winning-value,winning-price) distributions consistent with each prefix, so that the total number of distributions that we have to memoize in the course of the algorithm is exponentially large. To obtain a polynomially small cover we show that we can be coarse in our bookkeeping of the (winning-value, winning-price) distributions, without sacrificing much revenue. Indeed, it is exactly here where viewing our problem in the "upside-down" way illustrated in Figure 1 (i.e. targeting a cover of the output of circuit C rather than figuring out a sparse cover of the input) is important: we show that, as far as the expected revenue is concerned, we can discretize probabilities into multiples of $\frac{1}{(nr)^3}$ after each round of the DP without losing much revenue, and while keeping the size of the DP table from exploding. That the loss due to pruning the search space is not significant follows from a joint application of the coupling lemma and the optimal coupling theorem (see, e.g., [11]), after each step of the Dynamic Program.

7. The Algorithm for the Discrete Problem

In this section, we formalize our ideas from the previous section, providing our main algorithmic result. We assume that the pricing problem at hand is discrete: the value distributions are supported on a discrete set $\mathcal{S} = \{v^{(1)}, v^{(2)}, \ldots, v^{(k_1)}\}$, and the sought after price vector also lies in a discrete set $\left\{p^{(1)}, \ldots, p^{(k_2)}\right\}^n$, where both $\mathcal{S}$ and $\mathcal{P} := \left\{p^{(1)}, \ldots, p^{(k_2)}\right\}$ are given explicitly as part of the input, while our access to the value distributions may still be either explicit or via an oracle. We denote by OPT the optimal expected revenue for this problem, when the prices are restricted to set $\mathcal{P}$.

The Algorithm.: As a first step, we reduce our problem into a new one in polynomial-time, where additionally the probabilities that the value distributions assign to each point in $\mathcal{S}$ is an integer multiple of $1/(rn)^3$, where $r = \max\left\{\frac{p^{(j)}}{p^{(i)}}\right\}$. The loss in revenue from this reduction is at most an additive $\frac{4k_1}{rn^2}\min\{p^{(i)}\}$. Moreover, the construction is explicit, so from now on we can assume that we know the value distributions explicitly. Let us denote by $\{\hat{F}_i\}_i$ the rounded distributions and set $m := rn$ throughout this section.

The second phase of our algorithm is the Dynamic Program outlined in Section 6. We provide some further details on this next. Our program computes a Boolean function $g(i, \widehat{\text{Pr}})$, whose arguments lie in the following range: $i \in [n]$ and $\widehat{\text{Pr}} = (\widehat{\text{Pr}}_{1,1}, \widehat{\text{Pr}}_{1,2}, \ldots, \widehat{\text{Pr}}_{k_1,k_2})$, where each $\widehat{\text{Pr}}_{i_1,i_2} \in [0,1]$ is an integer multiple of $\frac{1}{m^3}$. The function g is stored in a table that has one cell for every setting of i and $\widehat{\Pr}$, and the cell contains a 0 or a 1 depending on the value of g at the corresponding input. In the terminology of the previous section, argument i indexes the last item in a prefix of the items and $\widehat{\Pr}$ is a (winning-value, winning-price) distribution in multiples of $\frac{1}{m^3}$. If $\widehat{\Pr}$ can arise from some pricing of the items $1 \dots i$ (up to discretization of probabilities into multiples of $\frac{1}{m^3}$), we intend to store $g(i, \widehat{\Pr}) = 1$; otherwise we store $g(i, \widehat{\Pr}) = 0$.

Due to lack of space we postpone the details of the Dynamic Program to the full paper.

Very briefly, the table is filled in a bottom-up fashion from $i = 1$ through n. At the end of the i-th iteration, we have computed all feasible "discretized" (winning-value,winning-price) distributions for the prefix $1 \dots i$, where "discretized" means that all probabilities have been rounded into multiples of $1/m^3$. For the next iteration, we try all possible prices $p^{(j)}$ for item $i+1$ and compute how each of the feasible discretized (winning-value,winning-price) distributions for the prefix $1 \dots i$ evolves into a discretized distribution for the prefix $1 \dots i+1$, setting the corresponding cell of layer $g(i+1, \cdot)$ of the DP table to 1. Notice, in particular, that *we lose accuracy in every step of the Dynamic Program*, as each step involves computing how a discretized distribution for items $1 \dots i$ evolves into a distribution for items $1 \dots i+1$ and then rounding the latter back again into multiples of $1/m^3$. We show in the analysis of our algorithm that the error accumulating from these roundings can be controlled via coupling arguments.

After computing g's table, we look at all cells such that $g(n, \widehat{\Pr}) = 1$ and evaluate the expected revenue resulting from the distribution $\widehat{\Pr}$, i.e.

$$\mathcal{R}_{\widehat{\Pr}} = \sum_{i_1 \in [k_1], i_2 \in [k_2]} p^{(i_2)} \cdot \widehat{\Pr}_{i_1, i_2} \cdot \mathbb{1}_{v^{(i_1)} \geq p^{(i_2)}}.$$

Having located the cell whose $\mathcal{R}_{\widehat{\Pr}}$ is the largest, we follow back-pointers to obtain a price vector consistent with $\widehat{\Pr}$. At some steps of the back-tracing, there may be multiple choices; we pick an arbitrary one to proceed.

Running Time and Correctness: We bound the algorithm's running time and revenue. Due to space limitations, we defer the proofs of the following lemmas to the full version.

Lemma 16. *Given an instance of* RESTRICTEDPRICE*, where the value distributions are supported on a discrete set $\mathcal{S}$ of cardinality k_1 and the prices are restricted to a discrete set $\mathcal{P}$ of cardinality k_2, the algorithm described in this section produces a price vector with expected revenue at least*

$$OPT - \left(\frac{2k_1k_2}{(nr)^2} + \frac{16}{n} \right) \cdot \min\{\mathcal{P}\},$$

where OPT is the optimal expected revenue, $\min\{\mathcal{P}\}$ is the lowest element of $\mathcal{P}$, and r is the ratio of the largest to the smallest element of $\mathcal{P}$.

Lemma 17. *The running time of the algorithm is polynomial in the size of the input and $(nr)^{O(k_1 k_2)}$.*

Intuitively, if we did not perform any rounding of distributions, our algorithm would have been *exact*, outputting an optimal price vector in $\{p^{(1)}, \dots, p^{(k_2)}\}^n$. What we show is that the roundings performed at the steps of the dynamic program are fine enough that do not become detrimental to the revenue. To show this, we use the probabilistic concepts of total variation distance and coupling of random variables, invoking the coupling lemma and the optimal coupling theorem after each step of the algorithm. This way, we show that the rounded (winning-value,winning-price) distributions maintained by the algorithm for each price vector are close in total variation distance to the corresponding exact distributions arising from these price vectors, culminating in Lemma 16.

Using Lemmas 16 and 17 and our work in the previous sections, we obtain our main algorithmic results in this paper (Theorems 1, 2, 3, and 4). Our analysis gives the following running times for the whole algorithm. For all $\epsilon > 0$, a $(1+\epsilon)$-factor approximation to the optimal revenue can be computed in time $n^{O(\frac{1}{\epsilon^7})}$ for MHR, and $n^{O(\frac{(\log(n))^9}{\epsilon^9})}$ for regular distributions. We only mildly tried to optimize the constants in our running times, and should be able to improve them with a more careful analysis.

Acknowledgments. We thank Jason Hartline for valuable feedback that helped improve the presentation.

REFERENCES

[1] S. Bhattacharya, G. Goel, S. Gollapudi and K. Munagala. Budget constrained auctions with heterogeneous items. *Proceedings of the ACM Symposium on Theory of Computing, STOC* 2010.

[2] L. Blumrosen and T. Holenstein. Posted prices vs. negotiations: an asymptotic analysis. *Proceedings of the ACM Conference on Electronic Commerce, EC* 2008.

[3] P. Briest, S. Chawla, R. Kleinberg and S. M. Weinberg. Pricing Randomized Allocations. *Proceedings of SODA* 2010.

[4] P. Briest. Uniform Budgets and the Envy-Free Pricing Problem. *Proceedings of ICALP* 2008.

[5] Y. Cai and C. Daskalakis. Extreme-Value Theorems for Optimal Multidimensional Pricing. *Arxiv Report*, 2011.

[6] S. Chawla, J. D. Hartline and R. D. Kleinberg. Algorithmic Pricing via Virtual Valuations. *Proceedings of the ACM Conference on Electronic Commerce,* EC 2007.

[7] S. Chawla, J. D. Hartline, D. Malec and B. Sivan. Multi-Parameter Mechanism Design and Sequential Posted Pricing. *Proceedings of the ACM Symposium on Theory of Computing, STOC* 2010.

[8] S. Chawla, D. Malec and B. Sivan. The Power of Randomness in Bayesian Optimal Mechanism Design. *Proceedings of the ACM Conference on Electronic Commerce, EC*, 2010.

[9] C. Daskalakis and S. M. Weinberg. On Optimal Multi-Dimensional Mechanism Design. *Manuscript,* 2011.

[10] L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer Series in Operations Research, 2006.

[11] R. Durrett. *Random Graph Dynamics*. Cambridge University Press, 2006.

[12] J. D. Hartline and V. Koltun. Near-Optimal Pricing in Near-Linear Time. *Proceedings of WADS,* 2005.

[13] A. M. Manelli and D. R. Vincent. Bundling as an Optimal Selling Mechanism for a Multiple-Good Monopolist. *Journal of Economic Theory,* 127(1):1–35, 2006.

[14] A. M. Manelli and D. R. Vincent. Multidimensional Mechanism Design: Revenue Maximization and the Multiple-Good Monopoly. *Journal of Economic Theory,* 137(1):153–185, 2007.

[15] R. B. Myerson. Optimal Auction Design. *Mathematics of Operations Research,* 1981.

[16] N. Nisan, T. Roughgarden, E. Tardos and V. V. Vazirani (eds.). *Algorithmic Game Theory*. Cambridge University Press, 2007.

[17] M. Pai and R. Vohra. Optimal auctions with financially constrained bidders. *Working Paper,* 2008.

[18] J. Thanassoulis. Haggling over substitutes. *Journal of Economic Theory,* 117(2):217245, 2004.

[19] R. B. Wilson. *Nonlinear Pricing.* Oxford University Press, 1997.

APPENDIX

1. Access to Value Distributions

We consider two ways that a distribution may be input to an algorithm.

Explicitly: In this case the distribution has to be discrete, and we are given its support as a list of numbers, and the probability that the distribution places on every point in the support. If a distribution is provided explicitly to an algorithm, the algorithm is said to be efficient, if it runs in time polynomial in the description complexity of the numbers required to specify the distribution.

As an Oracle: In this case, we are given an oracle that answers queries about the value of the cumulative distribution function on a queried point. In particular, a query to the oracle consists of a point x and a precision ϵ, and the oracle outputs a value of bit complexity polynomial in the description of x and ϵ, which is within ϵ from the value of the cumulative distribution function at point x. Moreover, we assume that we are given an *anchoring point* x^* such that the value of the cumulative distribution at that point is between two a priori known absolute constants c_1 and c_2, such that $0 < c_1 < c_2 < 1$. Having such a point is necessary, as otherwise it would be impossible to find any interesting point in the support of the distribution (i.e. any point where the cumulative is different than 0 or 1).

If a distribution is provided to an algorithm as an oracle, the algorithm is said to be efficient, if it runs in time polynomial in its other inputs and the bit complexity of x^*, ignoring the time spent by the oracle to answer queries (since this is not under the algorithm's control).

If we have a closed form formula for our input distribution, e.g. if our distribution is $\mathcal{N}(\mu, \sigma^2)$, we think of it as given to us as an oracle, answering queries of the form (x, ϵ) as specified above. In most common cases, such an oracle can be implemented so that it also runs efficiently in the description of the query.

2. Discussion of Theorem 14

In this section we play around with Theorem 14 to gain some intuition about its meaning:

• Suppose that we set all the t_i's equal to $t \geq 2n^2\alpha/\epsilon^2$. In this case, the homogenization property of Theorem 14 essentially states that the union bound is tight for t large enough. Indeed:

$$\begin{aligned}
&\Pr\left[\max_i\{X_{a_i}\} \geq t\right] \\
&\leq \left(\sum_{i=1}^{m} \Pr[X_{a_i} \geq t]\right) \\
&\leq \left(\frac{t - \frac{2\alpha}{\epsilon}}{t}\right) \cdot \Pr\left[\max_i\{X_{a_i}\} \geq t\right] \\
&\quad + \frac{7\epsilon}{tn} \cdot \left(\frac{2\alpha}{\epsilon} \cdot \Pr\left[\max_i\{X_{a_i}\} \geq \frac{2\alpha}{\epsilon}\right]\right) \\
&\leq \Pr\left[\max_i\{X_{a_i}\} \geq t\right] \\
&\quad + \frac{7\epsilon}{n} \cdot \left(\frac{2\alpha}{\epsilon} \cdot \Pr\left[\max_i\{X_{a_i}\} \geq \frac{2\alpha}{\epsilon}\right]\right).
\end{aligned}$$

This is not surprising, since for all i, the event $X_{a_i} \geq t$ only happens with tiny probability, by the anchoring property of α.

• Now let's try to set all the t_i's to the same value $t' > t \geq 2n^2\alpha/\epsilon^2$. The homogenization property can be used to obtain that the probability of the event $\max_i\{X_{a_i}\} \geq t'$ scales linearly in t'.

$$\begin{aligned}
&\Pr\left[\max_i\{X_{a_i}\} \geq t'\right] \\
&\leq \sum_{i=1}^{m} \Pr[X_{a_i} \geq t'] \\
&\leq \left(\frac{t - \frac{2\alpha}{\epsilon}}{t'}\right) \cdot \Pr\left[\max_i\{X_{a_i}\} \geq t\right] \\
&\quad + \frac{7\epsilon}{t'n} \cdot \left(\frac{2\alpha}{\epsilon} \cdot \Pr\left[\max_i\{X_{a_i}\} \geq \frac{2\alpha}{\epsilon}\right]\right) \\
&\leq \frac{1}{t'} \cdot \left[t \cdot \Pr\left[\max_i\{X_{a_i}\} \geq t\right]\right] \\
&\quad + \frac{7\epsilon}{t'n} \cdot \left(\frac{2\alpha}{\epsilon} \cdot \Pr\left[\max_i\{X_{a_i}\} \geq \frac{2\alpha}{\epsilon}\right]\right).
\end{aligned}$$

This follows easily from Markov's inequality, if the expression in the brackets is within a constant factor of $\mathbb{E}[\max_i\{X_{a_i}\}]$. The result is surprising as it is perfectly possible for that expression to be much smaller than $\mathbb{E}[\max_i\{X_{a_i}\}]$.

In the same spirit as the second point above, the theorem has interesting implications by setting the t_i's to different values.