

# CS264: Beyond Worst-Case Analysis

## Lecture #18: Pricing with an Unknown Distribution\*

Tim Roughgarden<sup>†</sup>

November 19, 2014

### 1 Preamble

This lecture continues last lecture's theme of designing an algorithm that performs well on inputs drawn from a fixed but a priori unknown distribution. Like last lecture, the design problem boils down to learning enough relevant statistics about the distribution to do almost as well as if the distribution were fully known.

### 2 A Pricing Problem

#### 2.1 Pricing with a Known Distribution

Consider the following extremely simple (but still interesting) problem. A seller has a single item or service that it wants to sell. There is a single potential buyer with an unknown willingness to pay — called a *valuation*  $v$  — for the item being sold. The seller chooses a price  $p$  and makes a take-it-or-leave offer of  $p$  to the buyer. If  $p \leq v$ , then the buyer accepts the offer and the seller makes revenue  $p$ . If  $p > v$ , then the buyer walks away and the seller earns nothing. If the seller knew  $v$ , it would of course just set  $p = v$  to maximize its revenue. With  $v$  unknown, what's a sensible choice of  $p$ ?

Intuitively, there is a trade-off as the seller increases the selling price  $p$ : the revenue of a sale goes up, but it is “less likely” that there will be a sale. With an unknown valuation and no further assumptions, it's not clear how to reason about this trade-off. But suppose the seller knows that the buyer's valuation is drawn from a distribution  $F$  — without knowing the actual realization  $v$ , of course — where  $F(x)$  denotes  $\Pr[v \leq x]$ .<sup>1</sup> Now, the expected

---

\*©2014, Tim Roughgarden.

<sup>†</sup>Department of Computer Science, Stanford University, 474 Gates Building, 353 Serra Mall, Stanford, CA 94305. Email: [tim@cs.stanford.edu](mailto:tim@cs.stanford.edu).

<sup>1</sup>For simplicity, we'll be sloppy about the difference between the quantities  $\Pr[v \leq x]$  and  $\Pr[v < x]$ . They are the same if  $F$  has no point masses. We leave it to the reader to generalize this lecture's results to distributions with point masses.

revenue (over  $v \sim F$ ) of a price  $p$  is just

$$\underbrace{p}_{\text{revenue of a sale}} \cdot \underbrace{(1 - F(p))}_{\text{chance of a sale}}.$$

The obvious strategy for a seller that knows  $F$  is to set  $p$  to maximize the expected revenue. A price  $p^*$  that accomplishes this is called a *monopoly price* for  $F$ , since it is the optimal price for a revenue-maximizing seller that is not facing any competition from other sellers.

For example, suppose that  $F$  is the uniform distribution on  $[0, 1]$ , with  $F(x) = x$  for  $x \in [0, 1]$ . Then, the expected revenue is  $p(1 - p)$ , which by calculus is maximized at  $p^* = \frac{1}{2}$  for an expected revenue of  $\frac{1}{4}$ . Alternatively, if  $F$  is the exponential distribution with rate 1, with  $F(x) = 1 - e^{-x}$ , then the expected revenue is  $pe^{-p}$ , which by calculus is maximized at  $p^* = 1$  for an expected revenue of  $\frac{1}{e}$ .

## 2.2 The Model

We now consider the case where the distribution  $F$  over the buyer's valuation is a priori unknown to the seller (but we continue to assume that it exists). Just like last lecture, we'll assume that the seller has partial knowledge about the distribution in the form of samples, and wants to use this information to do almost as well as if the distribution  $F$  were known. Formally, the input to the problem is  $m$  "training inputs"  $v_1, \dots, v_m$ , which are i.i.d. draws from  $F$ . The goal is to pick a price  $p = p(v_1, \dots, v_m)$ , as a function of the samples, so that

$$\underbrace{p \cdot (1 - F(p))}_{\text{seller's expected revenue}} \approx \underbrace{p^* \cdot (1 - F(p^*))}_{\text{optimal expected revenue}}, \quad (1)$$

where  $p^*$  is the monopoly price for  $F$ . The obvious interpretation of the samples  $v_1, \dots, v_m$  is as data from past transactions for comparable items with comparable types of buyers. In today's world, it's easy to imagine many ways in which an individual or company could acquire such information.

We next study the following question: how many samples  $m$  are necessary and sufficient to achieve (1), no matter what the valuation distribution  $F$  is? More formally, for the seller to achieve expected revenue at least  $1 - \epsilon$  times that of the monopoly price, how large does  $m$  need to be as a function of  $\epsilon$ ?<sup>2</sup>

## 2.3 An Impossibility Result

Without imposing any restriction on the distribution  $F$ , no positive result of this type is possible. Formally, for every fixed number  $m$  of samples, there exists a valuation distribution  $F$  such that a seller with  $m$  samples from  $F$  cannot obtain expected revenue close to optimal.

---

<sup>2</sup>For the seller's expected revenue, the expectation is over both the provided samples  $v_1, \dots, v_m$  (yesterday's buyers) and over the valuation of a new buyer that is offered a price of  $p(v_1, \dots, v_m)$  (tomorrow's buyer). For the (optimal) expected revenue achieved by the monopoly price, the expectation is only over the valuation of a (new) buyer.

To see this, consider distributions of the form  $\Pr[v = M] = \frac{1}{m^2}$  and  $\Pr[v = 0] = 1 - \frac{1}{m^2}$ , where  $M$  is an unknown large number, at least  $m^3$ . The monopoly price for such a distribution is obviously  $p^* = M$ , with expected revenue  $M/m^2$ . With high probability,  $m$  samples from such a distribution are all zero, leaving the seller to take a wild guess as to what  $M$  might be. Most of the time, such a guess will either be bigger than  $M$  (yielding zero revenue) or much smaller than  $M$  (yielding expected revenue much worse than optimal). We leave further details to the reader.

The point is that to rescue our goal of identifying a uniform (over  $F$ ) bound  $m$  on the number of samples needed to achieve a  $1 - \epsilon$  fraction of the optimal revenue, we need to impose additional restrictions on  $F$ . Happily, mild conditions are enough. We'll see two examples in this lecture.

## 2.4 A Positive Result for Bounded Valuations

For the rest of this section, we restrict attention to valuation distributions that have support in  $[1, H]$  with  $H < +\infty$ . Subject to this support constraint, the distribution  $F$  can be otherwise arbitrary. The main result of this section is an upper bound on the “sample complexity” of achieving a  $(1 - \epsilon)$ -approximation of the optimal expected revenue, as a function of  $H$  and  $\frac{1}{\epsilon}$ . This upper bound will follow from a sensible strategy for setting a price  $p$  as a function of the samples  $v_1, \dots, v_m$ .

There are two different ways to think about prices, and it's useful to have facility in switching back and forth between them. For simplicity, we'll assume from now on that the valuation distribution  $F$  has a finite and strictly positive density  $f$  on its support, so that  $F$  is continuous and strictly increasing, with a uniquely defined inverse  $F^{-1}$ . For a price  $p \geq 0$ , we define the *quantile*  $q(p) = 1 - F(p)$  as the probability of a sale at the price  $p$ . For every quantile  $q \in [0, 1]$ , we can define the corresponding price  $p(q) = F^{-1}(1 - q)$  — the price at which the selling probability is precisely  $q$ . Observe that  $p(q(p)) = p$  and  $q(p(q)) = q$  for all  $p$  and  $q$  — that is, the two functions are inverses. As the price  $p$  goes up, the corresponding quantile goes down; as the quantile  $q$  goes up, the corresponding price goes down. See Figure 1. The expected revenue can be written either as  $p \cdot q(p)$  or as  $q \cdot p(q)$ .

We follow a two-step approach to determining a good price from the samples. First, we identify a small set of statistics about the distribution  $F$  that, if known perfectly, are sufficient to obtain near-optimal revenue. Then we estimate these statistics using the samples. We introduced quantiles because they offer a convenient language in which to express a useful small set of statistics. The formal statement of the first step of our two-step approach is the following.

**Lemma 2.1** *Fix  $\epsilon > 0$  and a distribution  $F$  with support in  $[1, H]$ , and with continuous and positive density on its support. Define  $q_0 = \frac{1}{H}$ ,  $q_1 = \frac{1}{H}(1 + \epsilon)$ ,  $q_2 = \frac{1}{H}(1 + \epsilon)^2$ ,  $\dots$ ,  $q_s = 1$ , with  $s \approx \log_{1+\epsilon} H = \Theta(\epsilon^{-1} \log H)$ . Then,*

$$\max_{i=0}^s q_i \cdot p(q_i)$$

*is at least a  $1/(1 + \epsilon)$  fraction of the expected revenue of the monopoly price for  $F$ .*

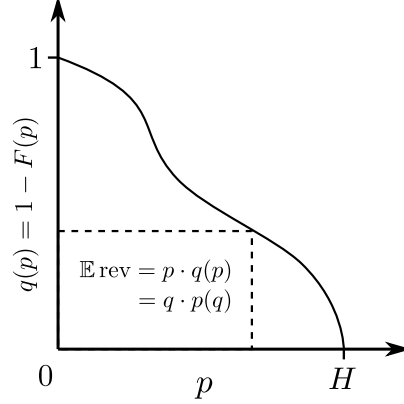


Figure 1: The quantile  $q$  and price  $p$  are inversely related, and for any fixed  $q$  (or fixed  $p$ ) the expected revenue is  $pq$ .

The point of Lemma 2.1 is that it reduces achieving near-optimal expected revenue to learning the  $q_i p(q_i)$ 's.

*Proof:* Let  $p^*$  denote the monopoly price for  $F$  and  $q^* = q(p^*)$  the corresponding quantile. First, observe that  $q^* \geq \frac{1}{H}$ : since valuations are guaranteed to lie in  $[1, H]$ , expected revenue 1 is trivial to achieve (just set  $p = 1$ ) and only prices that sell with probability at least  $1/H$  can do at least as well.

Let  $q_\ell$  denote the largest  $q_i$  bounded above by  $q^*$ . By definition of the  $q_i$ 's, we have  $q_\ell \geq q^*/(1 + \epsilon)$ . Since  $q_\ell \leq q^*$ , we also have  $p(q_\ell) \geq p(q^*) = p^*$ . Thus, the expected revenue of the quantile  $q_\ell$  is  $q_\ell \cdot p(q_\ell) \geq q^* \cdot p^*/(1 + \epsilon)$ , and the best of the  $q_i \cdot p(q_i)$ 's can only be better. ■

Given Lemma 2.1, a natural idea is to learn (approximations of) the  $q_i \cdot p(q_i)$ 's. Since there aren't too many of them, this seems like a reasonable goal. The good news is that Lemma 2.1 is robust under approximations, in multiple senses.

First, inspection of the proof of Lemma 2.1 shows that its conclusion remains valid for any set of quantiles satisfying the property that, for every  $t \in [\frac{1}{H}, 1]$ , there exists a quantile  $q_i$  in the set with  $q_i \in [\frac{t}{1+\epsilon}, t]$ .

Second, if we only have an estimate of each expected revenue  $q_i \cdot p(q_i)$  that is accurate to within  $1 \pm \epsilon$ , then the approximation guarantee in Lemma 2.1 degrades by at most an additional  $\frac{1-\epsilon}{1+\epsilon}$  factor. In the worst case, we underestimate the expected revenue of the best quantile by a  $1 - \epsilon$  factor, and overestimate that of the second-best quantile by a  $1 + \epsilon$  factor (see Figure 2). This loss doesn't bother us, since by resetting  $\epsilon$  to  $\epsilon/4$  at the beginning of the argument — increasing the sample complexity upper bound below by a constant factor — the approximation factor of  $(1 - \epsilon)/(1 + \epsilon)^2$  goes back to an approximation factor of  $1 - \epsilon$ .

With these two relaxations of Lemma 2.1, we're good to go. The algorithm is the following:

1. Receive as input samples  $v_1, \dots, v_m \sim F$ .

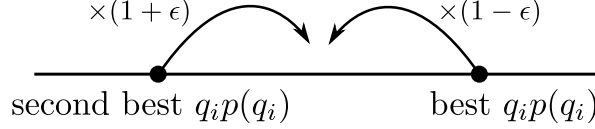


Figure 2: Having a  $1 \pm \epsilon$  estimate of the expected revenues degrades the approximation guarantee by only  $(1 - \epsilon)/(1 + \epsilon)$ .

2. Define “empirical quantiles” by

$$\hat{q}(v_i) = \frac{\text{number of } v_j\text{'s with } v_j \geq v_i}{m}.$$

[All empirical quantiles are multiples of  $\frac{1}{m}$ , with  $\frac{1}{m}$  for the maximum sample and 1 for the minimum sample.]

3. For the final offer price, use  $v_{i^*}$ , where

$$i^* \in \operatorname{argmax}_{i=1}^m v_i \cdot \hat{q}(v_i).$$

The following lemma states that, provided  $m$  is sufficiently large, the empirical quantiles used in the algorithm above satisfy the hypotheses of the relaxed version of Lemma 2.1.

**Lemma 2.2** *Fix an approximation parameter  $\epsilon > 0$  and assume that the number  $m$  of samples is at least  $c \frac{H}{\epsilon^2} \ln \frac{H}{\epsilon}$ , where  $c$  is a sufficiently large constant.*

(a) *With probability at least  $1 - \frac{\epsilon}{H}$ , for every  $t \in [\frac{1}{H}, 1]$ , there exists a sample  $v_i$  such that  $q(v_i) \in [\frac{t}{1+\epsilon}, t]$ .*

(b) *With probability at least  $1 - \frac{\epsilon}{H}$ , for every sample  $v_i$  with  $q(v_i) \geq \frac{1}{H}$ ,*

$$\hat{q}(v_i) \in [(1 - \epsilon)q(v_i), (1 + \epsilon)q(v_i)]$$

*and hence*

$$v_i \cdot \hat{q}(v_i) \in [(1 - \epsilon)v_i \cdot q(v_i), (1 + \epsilon)v_i \cdot q(v_i)].$$

(c) *With probability at least  $1 - \frac{\epsilon}{H}$ , the algorithm above will not return a sample  $v_i$  with  $q(v_i) < \frac{1}{H}$ .*

The proof of Lemma 2.2 is a nice exercise in applying Chernoff bounds, and we leave it to Homework #9.

The relaxed version of Lemma 2.1 and Lemma 2.2 together imply that, with high probability, our learning algorithm produces a price that achieves expected revenue at least  $(1 - \epsilon)^2/(1 + \epsilon)$  times that of a monopoly price. (Thus, the expected revenue earned by the algorithm is also close to that of the monopoly price.) In more detail, part (c) implies that

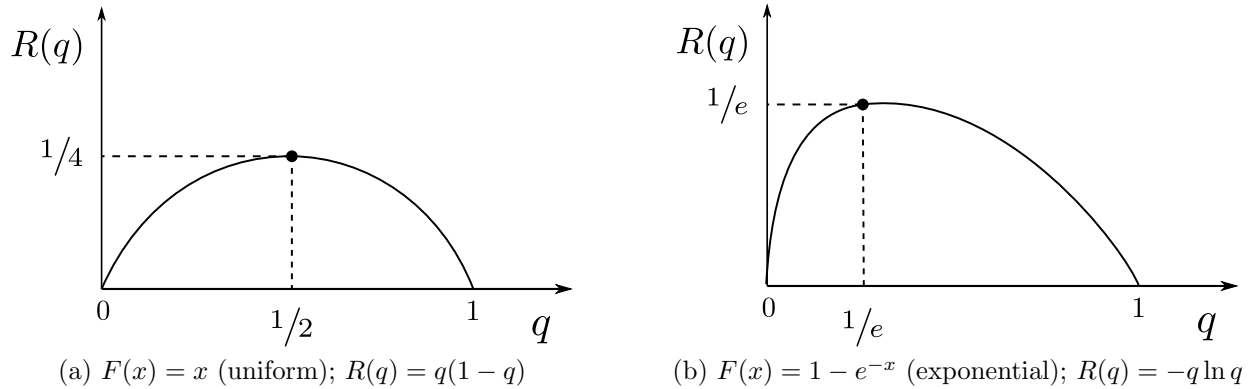


Figure 3: The revenue curve  $R(q)$  is defined as  $q \cdot p(q) = q \cdot F^{-1}(1 - q)$ . The highest point on the curve is the expected revenue of the monopoly price.

we can analyze our learning algorithm as if all of the quantiles  $q(v_i)$  lie between  $\frac{1}{H}$  and 1. Then, parts (a) and (b) of Lemma 2.2 are the hypotheses needed by the relaxed version of Lemma 2.1 to imply an approximation ratio of  $(1 - \epsilon)^2 / (1 + \epsilon)$ .

As a final remark, the sample complexity upper bound of  $O(\frac{H}{\epsilon^2} \log \frac{H}{\epsilon})$  is tight, up to the logarithmic factor [3].

### 3 The Single-Sample Regime

Last section we set a revenue target —  $1 - \epsilon$  times the expected revenue of the monopoly price — and asked how many samples were needed to achieve it. This section, we instead consider a limited number of samples and ask about the best-possible approximation guarantee that can be achieved with them. We restrict our attention to the single-sample regime ( $m = 1$ ) — where, amazingly, positive results are possible!

#### 3.1 A Guarantee for a Single Sample

Recall from last section the need for some kind of restriction on the valuation distribution  $F$ . Last section we restricted the support; here we'll use a different, but again reasonably mild, assumption. Stating the assumption requires defining the *revenue curve*  $R$  corresponding to a valuation distribution  $F$ , which plots the expected revenue  $R(q) = q \cdot p(q)$  as a function of the quantile  $q \in [0, 1]$ . See Figure 3. The highest point of the curve is the expected revenue of the monopoly price.

For example, if  $F$  is the uniform distribution on  $[0, 1]$ , then  $R(q) = q \cdot p(q) = q \cdot F^{-1}(1 - q) = q(1 - q)$ , which is the curve shown in Figure 3(a), with maximum expected revenue  $\frac{1}{4}$  at  $q = \frac{1}{2}$ . Alternatively, if  $F$  is the exponential distribution with rate 1, then  $R(q) = q \cdot p(q) = q \cdot F^{-1}(1 - q) = q \ln \frac{1}{q}$ , with maximum expected revenue  $\frac{1}{e}$  at  $q = \frac{1}{e}$ . See Figure 3(b).

We call a valuation distribution *regular* if its revenue curve is concave, with all of its

chords lying under the curve. The uniform and exponential distributions are examples, as are many other common distributions. Regularity is an incomparable assumption to bounded valuations — neither assumption implies the other.<sup>3</sup>

Now suppose we know nothing about a valuation distribution except for a single sample  $v$ , and have to set an offer price for a new buyer. What could we do? The first thing to try is to set the price  $p$  equal to  $v$ . This turns out to give a non-trivial revenue guarantee, assuming only that  $F$  is regular.<sup>4</sup> In other words, a random sample  $v$  from a regular valuation distribution  $F$  is a surprisingly good surrogate for the monopoly price of  $F$ .

**Theorem 3.1** ([1, 2]) *If  $F$  is a regular distribution and  $p \sim F$ , then*

$$\underbrace{\mathbf{E}_p[p \cdot (1 - F(p))]}_{\text{expected revenue of a random price}} \geq \underbrace{\frac{1}{2} \cdot p^*(1 - F(p^*))}_{\text{optimal revenue}},$$

where  $p^*$  is the monopoly price for  $F$ .

**Corollary 3.2** *With  $m = 1$  sample  $v$  from an unknown regular distribution  $F$ , setting  $p = v$  yields expected revenue at least 50% times that of the optimal price for  $F$ .*

*Proof of Theorem 3.1:* We give a proof by picture, see Figure 4. We can assume that  $p(1) = 0$  (why?), and hence the endpoints of the revenue curve are at  $(0, 0)$  and  $(1, 0)$ .

Let  $q^*$  denote the quantile of the monopoly price of  $F$ . The rectangle in Figure 4 has width 1 and height  $R(q^*)$ , and thus area  $R(q^*)$ , the optimal expected revenue.

Next, since  $p = F^{-1}(1 - q)$ , we have

$$\mathbf{E}_{p \sim F}[p \cdot (1 - F(p))] = \mathbf{E}_q \left[ \underbrace{F^{-1}(1 - q) \cdot q}_{\text{revenue } R(q) \text{ of } q} \right]. \quad (2)$$

Moreover, since  $p$  has distribution  $F$ ,  $F(p)$  is distributed uniformly over  $[0, 1]$ , so  $q = 1 - F(p)$  is also distributed uniformly over  $[0, 1]$ . Thus, the quantity in (2) is just the integral (with respect to the usual Lebesgue measure) of  $R(q)$  over  $[0, 1]$  — *the area under the revenue curve*.

We now use the regularity assumption. Since the revenue curve is concave by assumption, the triangle with endpoints  $(0, 0)$ ,  $(1, 0)$ , and  $(q^*, R(q^*))$  lies entirely under the revenue curve. The area under the revenue curve is at least that of this triangle (with base 1 and height  $R(q^*)$ ), which is  $\frac{1}{2} \cdot R(q^*)$  — 50% of the expected revenue earned by the monopoly price. ■

<sup>3</sup>The bad distributions of Section 2.3 are not regular. Multi-modal distributions, which can be interesting in applications, are also often not regular.

<sup>4</sup>Interestingly, there are settings where setting  $p = v$  is not the optimal way to use a single sample. Because of the asymmetry in regret between setting too high a price (which can lose all revenue) and setting too low a price (which loses only some of the potential revenue), it is sometimes better to scale  $v$  down by a constant factor [3].

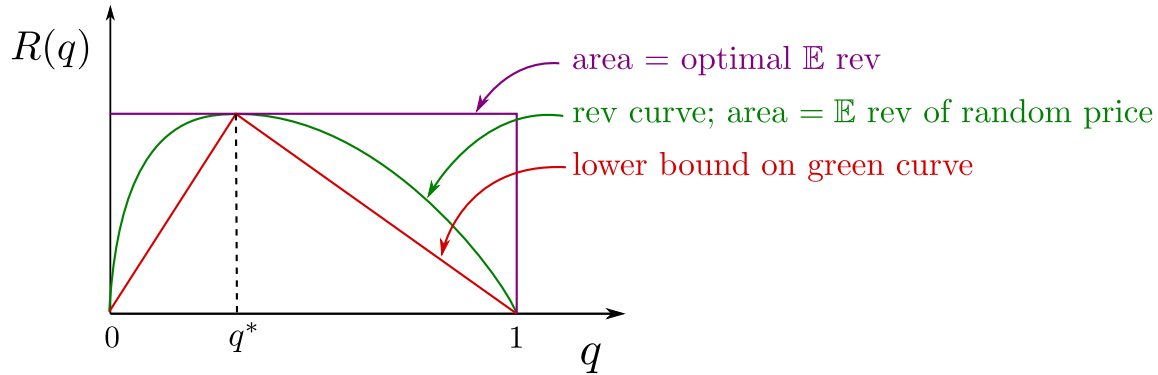


Figure 4: The orange triangle is a lower bound on the green revenue curve, since the green curve is concave by the regularity of  $F$ . Hence the area under the green curve is at least half the area of the purple rectangle, which is the optimal expected revenue.

### 3.2 Application to Auctions

We now generalize the pricing problem to an auction setting, where there are multiple potential buyers. Continue to assume a single seller of a single item or service, but now with  $n$  buyers, with valuations  $v_1, \dots, v_n$  drawn i.i.d. from a distribution  $F$ . The problem is to decide which buyer to sell to (if any), and at what price. If the valuations are unknown to the seller, a natural idea is to ask each buyer  $i$  for a bid  $b_i$ , and to sell the item to the highest bidder. But at what price? Your first instinct is probably to charge the winner their bid, known as a “first-price auction.” The *Vickrey* or *second-price* auction makes a different design choice.

#### Vickrey auction:

1. Collect a bid  $b_1, \dots, b_n$  from each bidder.
2. Award the item to the highest bidder  $i^* = \operatorname{argmax}_{i=1}^n b_i$ .
3. Charge the winner a price equal to the *second-highest* bid  $\max_{i \neq i^*} b_i$ .

Charging the second-highest bid seems less weird once you think about how things work on eBay, or at auction houses like Soethby’s. For example, on eBay, you submit a “maximum bid,” and eBay bids up on your behalf as needed (up to your maximum) to outbid the competition. Thus, if you win, you generally don’t pay your maximum bid — just whatever was needed to outbid everyone else, the second-highest bid (plus a small increment). Similarly, at an auction house, the winner is the last one with their hand up, and the final price is that at which the second-to-last bidder dropped out. The second-price auction is a sealed-bid (one-shot) version of this same idea. And it’s a very easy auction to participate in, as no bidder can ever gain anything by reporting a bid  $b_i$  different from their valuation  $v_i$  (see Homework #9). In game theory parlance, truthful bidding is a “dominant strategy,” and from now on we assume that all bidders bid their true valuations in a Vickrey auction.



The Vickrey auction is not designed to maximize the seller’s revenue. For example, if the highest bid is huge and the second-highest bid is tiny, the seller makes almost no revenue despite the fact that the winner is willing to pay a lot. Also note that the description of the second-price auction is not tailored to any distribution  $F$  over valuations — it is a totally general-purpose, prior-independent solution to the problem of selling a single item to buyers with unknown valuations.

Remarkably, in spite of these facts, the Vickrey auction *does* enjoy non-trivial revenue guarantees under reasonable assumptions. For example, the following guarantee is a consequence of Theorem 3.1; see Homework #9 for the proof.

**Corollary 3.3 ([1])** *For every regular distribution  $F$ , the expected revenue of the Vickrey auction with 2 bidders with i.i.d. valuations  $v_1, v_2 \sim F$  is at least  $p^* \cdot (1 - F(p^*))$ , where  $p^*$  is a monopoly price for  $F$ .*

That is, no matter what the regular distribution  $F$  is, the Vickrey auction with two bidders does at least as well as the optimal price for a single bidder. In fact, for every  $n \geq 1$ , the Vickrey auction with  $(n+1)$  i.i.d. bidders earns expected revenue at least that of the optimal selling procedure with merely  $n$  i.i.d. bidders from the same distribution [1]. This result — the *Bulow-Klemperer Theorem* — gives the very practical advice that, if you want to make money running an auction, it’s more important to increase competition, recruiting as many serious bidders as possible, than to gather detailed knowledge about bidders’ preferences (i.e. about  $F$ ) and then implement an auction tailored to these preferences. (Of course, do both if you can!) The Bulow-Klemperer theorem also implies that, with  $n$  bidders with valuations drawn i.i.d. from a regular distribution, the expected revenue of the Vickrey auction is at least  $\frac{n-1}{n}$  times that of any other auction (with the same number  $n$  of bidders) [4].

## References

- [1] J. Bulow and P. Klemperer. Auctions versus negotiations. *American Economic Review*, 86(1):180–194, 1996.
- [2] P. Dhangwatnotai, T. Roughgarden, and Q. Yan. Revenue maximization with a single sample. *Games and Economic Behavior*, 2015. To appear.
- [3] Z. Huang, Y. Mansour, and T. Roughgarden. Making the most of few samples. Working paper, 2014.
- [4] T. Roughgarden and M. Sundararajan. Is efficiency expensive? In *Third Workshop on Sponsored Search Auctions*, 2007.