Data-Driven Auction Design I
Model and Basic Techniques

Zhiyi Huang

University of Hong Kong

Single-Item Auction

- Sell 1 item to n bidders, to maximize revenue
- Bidder i's value v_i is drawn independently from D_i

Single-Item Auction

- Sell 1 item to n bidders, to maximize revenue
- Bidder i's value v_i is drawn independently from D_i
- Direct revelation auction
  1. Bidders bid $b_1, b_2, \ldots, b_n$
  2. Seller picks allocations $x_1, x_2, \ldots, x_n$ and payments $p_1, p_2, \ldots, p_n$
  3. Bidder i wins the item w.p. x_i, pays p_i, gets utility $v_i x_i - p_i$

Single-Item Auction

- Sell 1 item to n bidders, to maximize revenue
- Bidder i's value v_i is drawn independently from D_i
- Direct revelation auction
  1. Bidders bid $b_1, b_2, \ldots, b_n$
  2. Seller picks allocations $x_1, x_2, \ldots, x_n$ and payments $p_1, p_2, \ldots, p_n$
  3. Bidder i wins the item w.p. x_i, pays p_i, gets utility $v_i x_i - p_i$
- Dominant-Strategy Incentive Compatible (DSIC)

$$\forall i, v_i, b_i, b_{-i} : \quad v_i x_i(v_i, b_{-i}) - p_i(v_i, b_{-i}) \geq v_i x_i(b_i, b_{-i}) - p_i(b_i, b_{-i})$$

Single-Item Auction

- Sell 1 item to n bidders, to maximize revenue
- Bidder i's value v_i is drawn independently from D_i
- Direct revelation auction
  1. Bidders bid $b_1, b_2, \ldots, b_n$
  2. Seller picks allocations $x_1, x_2, \ldots, x_n$ and payments $p_1, p_2, \ldots, p_n$
  3. Bidder i wins the item w.p. x_i, pays p_i, gets utility $v_i x_i - p_i$
- Dominant-Strategy Incentive Compatible (DSIC)

$$\forall i, v_i, b_i, b_{-i} : \quad v_i x_i(v_i, b_{-i}) - p_i(v_i, b_{-i}) \geq v_i x_i(b_i, b_{-i}) - p_i(b_i, b_{-i})$$

- Individually Rational (IR)

$$\forall i, v_i, b_{-i} : \quad v_i x_i(v_i, b_{-i}) - p_i(v_i, b_{-i}) \geq 0$$

Myerson's Theory

- DSIC and IR are equivalent to
  1. $x_i(v_i, b_{-i})$ is monotone (e.g., step function)
  2. $p_i(v_i, b_{-i})$ is the area on the left of $x_i(v_i, b_{-i})$ as a function of v_i (e.g., threshold price above which $x_i = 1$, if x_i is a step function)

Myerson's Theory

- DSIC and IR are equivalent to
  1. $x_i(v_i, b_{-i})$ is monotone (e.g., step function)
  2. $p_i(v_i, b_{-i})$ is the area on the left of $x_i(v_i, b_{-i})$ as a function of v_i (e.g., threshold price above which $x_i = 1$, if x_i is a step function)
- Expected revenue is equivalent to expected virtual welfare

$$\mathbf{E} \sum_{i=1}^{n} \varphi_i(v_i) x_i(v)$$

where the virtual value φ_i is

$$\varphi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}$$

Myerson's Theory

- DSIC and IR are equivalent to
  1. $x_i(v_i, b_{-i})$ is monotone (e.g., step function)
  2. $p_i(v_i, b_{-i})$ is the area on the left of $x_i(v_i, b_{-i})$ as a function of v_i (e.g., threshold price above which $x_i = 1$, if x_i is a step function)
- Expected revenue is equivalent to expected virtual welfare

$$\mathbf{E} \sum_{i=1}^{n} \varphi_i(v_i) x_i(v)$$

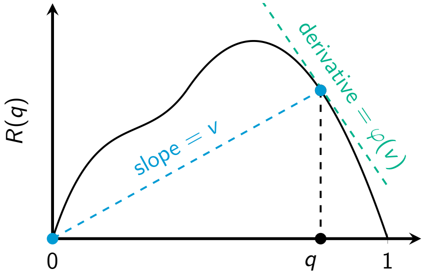
where the virtual value φ_i is

$$\varphi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}$$

- Myerson's optimal auction deferred to next lecture

Optimal Pricing in the Single-Bidder Case

- Sell 1 item to 1 bidder, whose value v is drawn from D
- Every DSIC and IR auction is equivalent to posting a price p
- Revenue of price p is $p \cdot q(p)$, where $q(p) = 1 - F(p)$ is p's quantile
- Revenue curve in quantile space $R(q) = v(q) \cdot q$



Data-Driven Optimal Pricing

Optimal pricing is easy...

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from D as input
  - Output a price p

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from D as input
  - Output a price p
- How many samples are needed to pick a near optimal p "up to an ε margin"?

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from D as input
  - Output a price p
- How many samples are needed to pick a near optimal p "up to an ε margin"?
  - ε additive approximation

    $[0, 1]$-bounded distributions **(illustrative example)**

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from D as input
  - Output a price p
- How many samples are needed to pick a near optimal p "up to an ε margin"?
  - ε additive approximation

    $[0, 1]$-bounded distributions **(illustrative example)**
  - $1 - \varepsilon$ (multiplicative) approximation

    Regular distributions (i.e., concave revenue curve)

    MHR distributions (i.e., "strongly concave" revenue curve)

    $[1, H]$-bounded distributions

Data-Driven Optimal Pricing

Optimal pricing is easy... but how much information is needed?

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from D as input
  - Output a price p
- How many samples are needed to pick a near optimal p "up to an ε margin"?
  - ε additive approximation

    $[0, 1]$-bounded distributions **(illustrative example)**
  - $1 - \varepsilon$ (multiplicative) approximation

    Regular distributions (i.e., concave revenue curve)

    MHR distributions (i.e., "strongly concave" revenue curve)

    $[1, H]$-bounded distributions
- The sample complexity is smallest number of samples needed

Model

Basic Techniques

Upper Bound Techniques

Lower Bound Techniques

Settling the Single-Item Single-Bidder Case

Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

- Consider a $[0, 1]$-bounded distribution, and ε additive approximation
- **Plan:** Estimate the revenue of every price up to ε additive error

Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

- Consider a $[0,1]$-bounded distribution, and ε additive approximation
- **Plan:** Estimate the revenue of every price up to ε additive error
    1. Estimate the revenue of one price p up to ε

Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

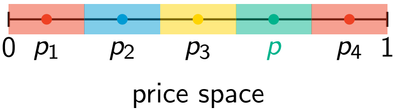
- Consider a $[0, 1]$-bounded distribution, and ε additive approximation
- **Plan:** Estimate the revenue of every price up to ε additive error
  1. Estimate the revenue of one price p up to ε
  2. Prices "close to" p cannot yield much higher revenue (up to ε)

Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

- Consider a $[0, 1]$-bounded distribution, and ε additive approximation
- **Plan:** Estimate the revenue of every price up to ε additive error
  1. Estimate the revenue of one price p up to ε
  2. Prices "close to" p cannot yield much higher revenue (up to ε)

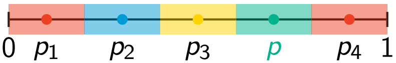
     $\Rightarrow$ Consider finitely many prices whose "neighborhoods" cover $[0, 1]$



Basic Upper Bound Techniques

Empirical Revenue Maximizer (ERM). Choose the price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

- Consider a $[0,1]$-bounded distribution, and ε additive approximation
- **Plan:** Estimate the revenue of every price up to ε additive error
  1. Estimate the revenue of one price p up to ε
  2. Prices "close to" p cannot yield much higher revenue (up to ε)
     $\Rightarrow$ Consider finitely many prices whose "neighborhoods" cover $[0,1]$
  3. Estimate the revenue of all these representative prices up to ε



price space

Step 1: Estimate the Revenue of One Price p

It suffices to estimate quantile $q(p) = 1 - F(p)$

Step 1: Estimate the Revenue of One Price p

It suffices to estimate quantile $q(p) = 1 - F(p)$

Theorem (Chernoff-Hoeffding, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ *are i.i.d. RV over* $[0, 1]$. *Let* $\mu = \mathbf{E}\, X_i$. *With probability* $1 - \delta$ *we have*

$$\left| \frac{1}{m} \sum_{i=1}^{m} X_i - \mu \right| \lesssim \sqrt{\frac{\log \frac{1}{\delta}}{m}}$$

Step 1: Estimate the Revenue of One Price p

It suffices to estimate quantile $q(p) = 1 - F(p)$

Theorem (Chernoff-Hoeffding, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ are i.i.d. RV over $[0,1]$. Let $\mu = \mathbf{E}\, X_i$. With probability $1-\delta$ we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \lesssim \sqrt{\frac{\log \frac{1}{\delta}}{m}}$$

Conclusion: Using $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples $v_1, v_2, \ldots, v_m \overset{\text{i.i.d.}}{\sim} D$ and letting $X_i = \mathbf{1}_{v_i \geq p}$, we can estimate $q(p)$ (and thus p's revenue) up to ε additive error w.p. $1-\delta$

Step 2: Covering the Price Space

Consider $\tilde{p}$ that is "close to" p. Can $\tilde{p}$'s revenue be much larger than p's?

$$\tilde{p} \cdot q(\tilde{p}) \quad \text{v.s.} \quad p \cdot q(p)$$

Step 2: Covering the Price Space

Consider $\tilde{p}$ that is "close to" p. Can $\tilde{p}$'s revenue be much larger than p's?

$$\tilde{p} \cdot q(\tilde{p}) \quad \text{v.s.} \quad p \cdot q(p)$$

1. If $p + \varepsilon \geq \tilde{p} > p$, then:

$$\begin{aligned} \tilde{p} \cdot q(\tilde{p}) &\leq \tilde{p} \cdot q(p) \\ &\leq (p + \varepsilon) \cdot q(p) \\ &\leq p \cdot q(p) + \varepsilon \end{aligned}$$

Step 2: Covering the Price Space

Consider $\tilde{p}$ that is "close to" p. Can $\tilde{p}$'s revenue be much larger than p's?

$$\tilde{p} \cdot q(\tilde{p}) \quad \text{v.s.} \quad p \cdot q(p)$$

1. If $p + \varepsilon \geq \tilde{p} > p$, then:

$$\begin{aligned} \tilde{p} \cdot q(\tilde{p}) &\leq \tilde{p} \cdot q(p) \\ &\leq \big(p + \varepsilon\big) \cdot q(p) \\ &\leq p \cdot q(p) + \varepsilon \end{aligned}$$

2. If $p > \tilde{p} \geq p - \varepsilon$, then $\tilde{p} \cdot q(\tilde{p})$ could be almost $p \cdot q(p)$
   e.g., $p = 1$, $\tilde{p} = 0.98$, and D is point mass at 0.99

Step 2: Covering the Price Space

Consider $\tilde{p}$ that is "close to" p. Can $\tilde{p}$'s revenue be much larger than p's?

$$\tilde{p} \cdot q(\tilde{p}) \quad \text{v.s.} \quad p \cdot q(p)$$

1. If $p + \varepsilon \geq \tilde{p} > p$, then:

$$\begin{aligned} \tilde{p} \cdot q(\tilde{p}) &\leq \tilde{p} \cdot q(p) \\ &\leq \big(p + \varepsilon\big) \cdot q(p) \\ &\leq p \cdot q(p) + \varepsilon \end{aligned}$$

2. If $p > \tilde{p} \geq p - \varepsilon$, then $\tilde{p} \cdot q(\tilde{p})$ could be almost $p \cdot q(p)$
   e.g., $p = 1$, $\tilde{p} = 0.98$, and D is point mass at 0.99

Conclusion: p covers $[p, p + \varepsilon]$; prices $0, \varepsilon, 2\varepsilon, \ldots, 1 - \varepsilon$ cover the price space $[0, 1]$

Step 3: Estimate Revenue of All Representative Prices

- Using $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, we can estimate $q(p)$ (and thus p's revenue) up to ε additive error w.p. $1 - \delta$
- Prices $0, \varepsilon, 2\varepsilon, \ldots, 1 - \varepsilon$ cover the price space $[0, 1]$

Step 3: Estimate Revenue of All Representative Prices

- Using $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, we can estimate $q(p)$ (and thus p's revenue) up to ε additive error w.p. $1 - \delta$
- Prices $0, \varepsilon, 2\varepsilon, \ldots, 1 - \varepsilon$ cover the price space $[0, 1]$

Theorem (Union Bound)

For any (bad) events $E_1, E_2, \ldots, E_n$, we have $\mathbf{Pr}\left[E_1 \cup E_2 \cup \cdots \cup E_n\right] \leq \sum_{i=1}^{n} \mathbf{Pr}\left[E_i\right]$

Step 3: Estimate Revenue of All Representative Prices

- Using $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, we can estimate $q(p)$ (and thus p's revenue) up to ε additive error w.p. $1 - \delta$
- Prices $0, \varepsilon, 2\varepsilon, \ldots, 1 - \varepsilon$ cover the price space $[0, 1]$

Theorem (Union Bound)

For any (bad) events $E_1, E_2, \ldots, E_n$, *we have* $\mathbf{Pr}\left[E_1 \cup E_2 \cup \cdots \cup E_n\right] \leq \sum_{i=1}^{n} \mathbf{Pr}\left[E_i\right]$

- If we estimate each representative price's revenue up to ε w.p. $1 - \varepsilon\delta$, then we estimate all of them w.p. at least $1 - \delta$

Step 3: Estimate Revenue of All Representative Prices

- Using $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, we can estimate $q(p)$ (and thus p's revenue) up to ε additive error w.p. $1 - \delta$
- Prices $0, \varepsilon, 2\varepsilon, \ldots, 1 - \varepsilon$ cover the price space $[0, 1]$

Theorem (Union Bound)

For any (bad) events $E_1, E_2, \ldots, E_n$, we have $\mathbf{Pr}\big[E_1 \cup E_2 \cup \cdots \cup E_n\big] \leq \sum_{i=1}^{n} \mathbf{Pr}\big[E_i\big]$

- If we estimate each representative price's revenue up to ε w.p. $1 - \varepsilon\delta$, then we estimate all of them w.p. at least $1 - \delta$

Conclusion: Using $m \gtrsim \frac{\log \frac{1}{\varepsilon\delta}}{\varepsilon^2}$ i.i.d. samples, we can estimate the revenue of all prices up to ε additive error w.p. $1 - \delta$

Upper Bound for $[0, 1]$-Bounded Distribution

Empirical Revenue Maximizer (ERM). Return price p that maximizes revenue w.r.t. uniform distribution over the samples (empirical distribution).

Theorem

ERM using $m \gtrsim \frac{\log \frac{1}{\varepsilon\delta}}{\varepsilon^2}$ *samples is an* ε *additiive approximation w.p.* $1 - \delta$*.*

Le Cam's Method (a.k.a., the Two-Point Method)

- Consider two value distributions P and Q that are
  1. Sufficiently "similar"
     One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q, say, w.p. $\frac{2}{3}$

Le Cam's Method (a.k.a., the Two-Point Method)

- Consider two value distributions P and Q that are

  1. Sufficiently "similar"
     One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q, say, w.p. $\frac{2}{3}$
  2. Sufficiently "different"
     No price p is an ε additive approximation for both P and Q

Le Cam's Method (a.k.a., the Two-Point Method)

- Consider two value distributions P and Q that are
  1. Sufficiently "similar"
     One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q, say, w.p. $\frac{2}{3}$
  2. Sufficiently "different"
     No price p is an ε additive approximation for both P and Q

- We next present
  1. Statistical distances that characterize the number of samples needed to distinguish two distributions
  2. Sufficient condition under which two distributions are "similar" enough
  3. Construction of P and Q

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$?

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$? What if $s = c$?

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$? What if $s = c$?
- Following this strategy, what is the total error?

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right]$$

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$? What if $s = c$?
- Following this strategy, what is the total error?

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right]$$

- Total variation distance

$$\mathrm{TV}(P, Q) = \frac{1}{2}\left\|P - Q\right\|_1$$

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$? What if $s = c$?
- Following this strategy, what is the total error?

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right]$$

- Total variation distance

$$\mathrm{TV}(P, Q) = \frac{1}{2}\left\|P - Q\right\|_1 = \frac{1}{2}\sum_{v}\left|P(v) - Q(v)\right|$$

Distinguish P and Q with One Sample

	a	b	c	d
P	0.1	0.2	0.3	0.4
Q	0.4	0.3	0.2	0.1

Suppose that $D \in \{P, Q\}$ and you draw one sample $s \sim D$

- If $s = a$, would you predict $D = P$ or $D = Q$? What if $s = c$?
- Following this strategy, what is the total error?

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right] = 1 - \mathrm{TV}(P, Q)$$

- Total variation distance

$$\mathrm{TV}(P, Q) = \frac{1}{2}\left\|P - Q\right\|_1 = \frac{1}{2}\sum_v \left|P(v) - Q(v)\right|$$

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m
- Minimum total error for distinguishing P and Q with m samples is:

$$1 - \mathrm{TV}(P^m, Q^m)$$

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m
- Minimum total error for distinguishing P and Q with m samples is:

$$1 - \mathrm{TV}(P^m, Q^m)$$

- Hard to reason about $\mathrm{TV}(P^m, Q^m)$ directly... We need a manageable proxy

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m
- Minimum total error for distinguishing P and Q with m samples is:

$$1 - \mathrm{TV}(P^m, Q^m)$$

- Hard to reason about $\mathrm{TV}(P^m, Q^m)$ directly... We need a manageable proxy
- Kullback-Leibler (KL) divergence

$$\mathrm{KL}(P \| Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$$

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m
- Minimum total error for distinguishing P and Q with m samples is:

$$1 - \mathrm{TV}(P^m, Q^m)$$

- Hard to reason about $\mathrm{TV}(P^m, Q^m)$ directly... We need a manageable proxy
- Kullback-Leibler (KL) divergence

$$\mathrm{KL}(P \| Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$$

  - Relation to TV (Pinsker's inequality) $\qquad \mathrm{TV}(P, Q) \leq \sqrt{\frac{1}{2}\mathrm{KL}(P \| Q)}$

Distinguish P and Q with Multiple Samples

- **Observation:** m samples from D $\Leftrightarrow$ one sample from D^m
- Minimum total error for distinguishing P and Q with m samples is:

$$1 - \mathrm{TV}(P^m, Q^m)$$

- Hard to reason about $\mathrm{TV}(P^m, Q^m)$ directly... We need a manageable proxy
- Kullback-Leibler (KL) divergence

$$\mathrm{KL}(P \| Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$$

  - Relation to TV (Pinsker's inequality) $\quad \mathrm{TV}(P, Q) \leq \sqrt{\frac{1}{2}\mathrm{KL}(P \| Q)}$
  - Direct sum $\quad \mathrm{KL}(P^m \| Q^m) = m \cdot \mathrm{KL}(P \| Q)$

Distinguish P and Q with Samples: a Summary

- **What we want:** One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q w.p. $\frac{2}{3}$

Distinguish P and Q with Samples: a Summary

- **What we want:** One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q w.p. $\frac{2}{3}$
- **Contrapositive:** If we have less than $m \precsim \frac{1}{\varepsilon^2}$ samples, then

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right] > \frac{2}{3}$$

Distinguish P and Q with Samples: a Summary

- **What we want:** One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q w.p. $\frac{2}{3}$
- **Contrapositive:** If we have less than $m \eqsim \frac{1}{\varepsilon^2}$ samples, then

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right] > \frac{2}{3}$$

- **Characterization via TV:**

$$\mathrm{TV}(P^m, Q^m) \leq \frac{1}{3}$$

Distinguish P and Q with Samples: a Summary

- **What we want:** One needs $m \gtrsim \frac{1}{\varepsilon^2}$ samples to distinguish P and Q w.p. $\frac{2}{3}$
- **Contrapositive:** If we have less than $m \eqsim \frac{1}{\varepsilon^2}$ samples, then

$$\mathbf{Pr}\left[\text{predict } P \mid D = Q\right] + \mathbf{Pr}\left[\text{predict } Q \mid D = P\right] > \frac{2}{3}$$

- **Characterization via TV:**

$$\mathrm{TV}(P^m, Q^m) \leq \frac{1}{3}$$

- **Characterization via KL:**

$$\mathrm{KL}(P^m \| Q^m) \lesssim 1 \quad \Rightarrow \quad \mathrm{KL}(P \| Q) \lesssim \frac{1}{m} \eqsim \varepsilon^2$$

Sufficient Condition for $\mathrm{KL}(P\|Q) \leq \varepsilon^2$

Reminder
$\mathrm{KL}(P\|Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$

Lemma
Suppose that $e^{-\varepsilon} \leq \frac{P(v)}{Q(v)} \leq e^{\varepsilon}$ *for any* v*. We have:*

$$\mathrm{KL}(P\|Q) \lesssim \varepsilon^2$$

Sufficient Condition for $\mathrm{KL}(P\|Q) \le \varepsilon^2$

Reminder
$\mathrm{KL}(P\|Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$

Lemma

Suppose that $e^{-\varepsilon} \le \frac{P(v)}{Q(v)} \le e^{\varepsilon}$ for any v. We have:

$$\mathrm{KL}(P\|Q) \lesssim \varepsilon^2$$

$$\mathrm{KL}(P\|Q) \le \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|P)$$

Sufficient Condition for $\mathrm{KL}(P\|Q) \le \varepsilon^2$

Reminder
$\mathrm{KL}(P\|Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$

Lemma
Suppose that $e^{-\varepsilon} \le \frac{P(v)}{Q(v)} \le e^{\varepsilon}$ for any v. We have:

$$\mathrm{KL}(P\|Q) \lesssim \varepsilon^2$$

$$\mathrm{KL}(P\|Q) \le \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|P)$$

$$= \sum_v (P(v) - Q(v)) \log \frac{P(v)}{Q(v)}$$

Sufficient Condition for $\mathrm{KL}(P\|Q) \le \varepsilon^2$

Reminder
$\mathrm{KL}(P\|Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$

Lemma
Suppose that $e^{-\varepsilon} \le \frac{P(v)}{Q(v)} \le e^{\varepsilon}$ for any v. We have:

$$\mathrm{KL}(P\|Q) \lesssim \varepsilon^2$$

$$\begin{aligned}
\mathrm{KL}(P\|Q) &\le \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|P) \\
&= \sum_v (P(v) - Q(v)) \log \frac{P(v)}{Q(v)} \\
&\le \sum_v (e^{\varepsilon} - 1) \min\{P(v), Q(v)\} \cdot \varepsilon
\end{aligned}$$

Sufficient Condition for $\mathrm{KL}(P\|Q) \le \varepsilon^2$

Reminder
$\mathrm{KL}(P\|Q) = \sum_v P(v) \log \frac{P(v)}{Q(v)}$

Lemma
Suppose that $e^{-\varepsilon} \le \frac{P(v)}{Q(v)} \le e^{\varepsilon}$ for any v. We have:

$$\mathrm{KL}(P\|Q) \lesssim \varepsilon^2$$

$$\begin{aligned}
\mathrm{KL}(P\|Q) &\le \mathrm{KL}(P\|Q) + \mathrm{KL}(Q\|P) \\
&= \sum_v (P(v) - Q(v)) \log \frac{P(v)}{Q(v)} \\
&\le \sum_v (e^{\varepsilon} - 1) \min\{P(v), Q(v)\} \cdot \varepsilon \\
&\le (e^{\varepsilon} - 1)\varepsilon
\end{aligned}$$

Lower Bound for $[0,1]$-Bounded Distributions

Theorem

Any ε additive approximation algorithm uses at least $m \gtrsim \frac{1}{\varepsilon^2}$ samples.

- Construct two $[0,1]$-bounded value distributions P and Q that are
  1. "Similar": For any v, $e^{-\varepsilon} \leq \frac{P(v)}{Q(v)} \leq e^{\varepsilon}$
  2. "Different": No price p is an ε additive approximation for both P and Q

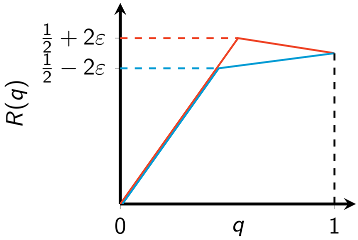
Lower Bound for $[0,1]$-Bounded Distributions

Theorem

Any ε additive approximation algorithm uses at least $m \gtrsim \frac{1}{\varepsilon^2}$ samples.

- Construct two $[0,1]$-bounded value distributions P and Q that are
  1. "Similar": For any v, $e^{-\varepsilon} \le \frac{P(v)}{Q(v)} \le e^{\varepsilon}$
  2. "Different": No price p is an ε additive approximation for both P and Q

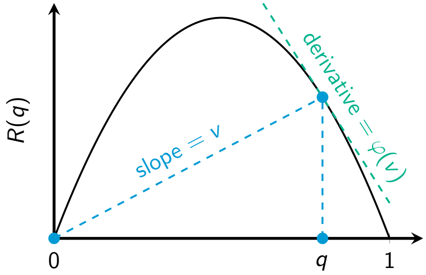
v	$\frac{1}{2}$	1
$P(v)$	$\frac{1}{2}+2\varepsilon$	$\frac{1}{2}-2\varepsilon$
$Q(v)$	$\frac{1}{2}-2\varepsilon$	$\frac{1}{2}+2\varepsilon$



Distributions	Sample Complexity
$[0, 1]$-Bounded	$\frac{1}{\varepsilon^2}$
Regular distributions	
MHR distributions	
$[1, H]$-bounded distributions	

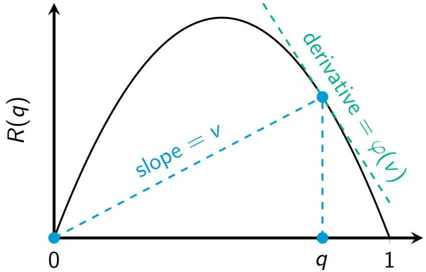
Regular Distributions

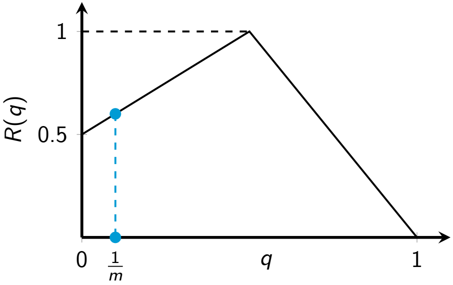
- Value distribution D is regular if $\varphi_D(v)$ is nondecreasing
  $\Leftrightarrow$ The revenue curve $R(q)$ is concave



Regular Distributions

- Value distribution D is regular if $\varphi_D(v)$ is nondecreasing
  $\Leftrightarrow$ The revenue curve $R(q)$ is concave





- ERM does not converge for some regular distribution
  - With constant probability we get two samples with quantiles less than $\frac{1}{m}$

What goes wrong?

1. Estimate the revenue of one price p up to $1 - \varepsilon \approx e^{-\varepsilon}$ approximation
2. Prices between p and $e^{\varepsilon} p$ cannot yield much higher revenue

   $\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

3. Estimate the revenue of all these representative prices

What goes wrong?

1. Estimate the revenue of one price p up to $1-\varepsilon \approx e^{-\varepsilon}$ approximation

Theorem (Bernstein Inequality, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ are i.i.d. RV over $[0,1]$. Let $\mu = \mathbf{E}\, X_i$. With probability $1-\delta$ we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \lesssim \max\left\{ \sqrt{\frac{\mu(1-\mu)\log\frac{1}{\delta}}{m}}, \frac{\log\frac{1}{\delta}}{m} \right\}$$

What goes wrong?

1. Estimate the revenue of one price p up to $1 - \varepsilon \approx e^{-\varepsilon}$ approximation

Theorem (Bernstein Inequality, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ are i.i.d. RV over $[0,1]$. Let $\mu = \mathbf{E}\, X_i$. With probability $1-\delta$ we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \lesssim \max\left\{\sqrt{\frac{\mu(1-\mu)\log\frac{1}{\delta}}{m}}, \frac{\log\frac{1}{\delta}}{m}\right\}$$

What goes wrong?

1. Estimate the revenue of one price p up to $1-\varepsilon \approx e^{-\varepsilon}$ approximation
   - To get $\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \leq \varepsilon\mu$ we need $m \gtrsim \frac{\log\frac{1}{\delta}}{\mu\varepsilon^2}$ samples

Theorem (Bernstein Inequality, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ are i.i.d. RV over $[0,1]$. Let $\mu = \mathbf{E}\, X_i$. With probability $1-\delta$ we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \lesssim \max\left\{\sqrt{\frac{\mu(1-\mu)\log\frac{1}{\delta}}{m}}, \frac{\log\frac{1}{\delta}}{m}\right\}$$

What goes wrong?

1. Estimate the revenue of one price p up to $1-\varepsilon \approx e^{-\varepsilon}$ approximation
   - To get $\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \leq \varepsilon\mu$ we need $m \gtrsim \frac{\log\frac{1}{\delta}}{\mu\varepsilon^2}$ samples
   - **Unbounded** when for small quantile μ (i.e., high prices)

Theorem (Bernstein Inequality, User-Friendly Version)

$X_1, X_2, \ldots, X_m$ are i.i.d. RV over $[0,1]$. Let $\mu = \mathbf{E}\, X_i$. With probability $1-\delta$ we have

$$\left|\frac{1}{m}\sum_{i=1}^{m} X_i - \mu\right| \lesssim \max\left\{\sqrt{\frac{\mu(1-\mu)\log\frac{1}{\delta}}{m}}, \frac{\log\frac{1}{\delta}}{m}\right\}$$

What goes wrong?

2. Prices between p and $e^{\varepsilon} p$ cannot yield much higher revenue

$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

What goes wrong?

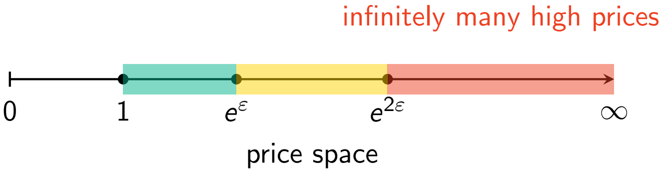
2. Prices between p and $e^{\varepsilon}p$ cannot yield much higher revenue

$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

What goes wrong?

2. Prices between p and $e^{\varepsilon}p$ cannot yield much higher revenue

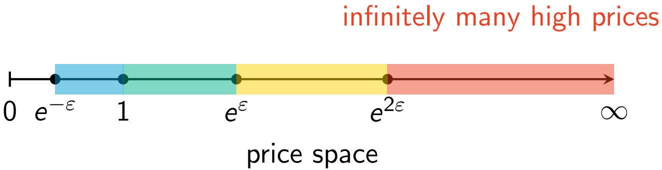
$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$



What goes wrong?

2. Prices between p and $e^{\varepsilon}p$ cannot yield much higher revenue

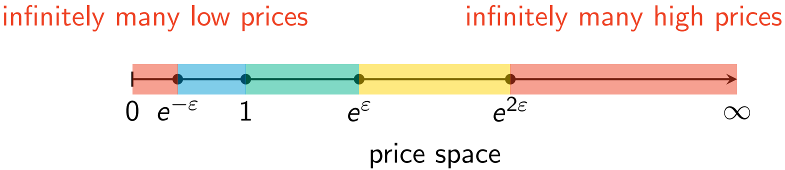
$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$



What goes wrong?

2. Prices between p and $e^{\varepsilon} p$ cannot yield much higher revenue

$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

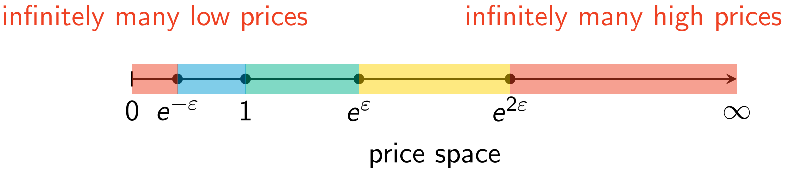


What goes wrong?

2. Prices between p and $e^{\varepsilon}p$ cannot yield much higher revenue

$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

- "Extremely low" prices are not relevant anyway

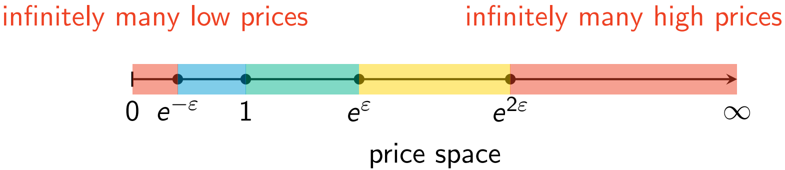


What goes wrong?

2. Prices between p and $e^{\varepsilon}p$ cannot yield much higher revenue

$\Rightarrow$ Consider finitely(?) many prices whose "neighborhoods" cover $[0, \infty)$

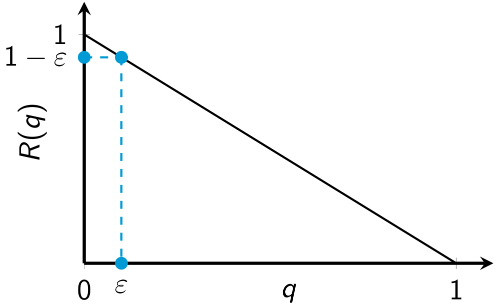
- "Extremely low" prices are not relevant anyway
- "Extremely high" prices will be "truncated" algorithimically



Existence of a "Good Enough" Price with "Large" Quantile

Observation: By concavity of revenue curve, there exists a price p such that

1. It is an $1-\varepsilon$ approximation
2. Its quantile is at least ε



Upper Bound for Regular Distributions

q-**Guarded ERM.** Return price p that maximizes the empirical revenue, among prices whose empirical quantiles are at least q.

Upper Bound for Regular Distributions

q-Guarded ERM. Return price p that maximizes the empirical revenue, among prices whose empirical quantiles are at least q.

Theorem

ε-Guarded ERM using $m \gtrsim \frac{\log \frac{1}{\varepsilon\delta}}{\varepsilon^3}$ samples is an $1-\varepsilon$ approximation w.p. $1-\delta$.

Upper Bound for Regular Distributions

> ***q*-Guarded ERM.** Return price p that maximizes the empirical revenue, among prices whose empirical quantiles are at least q.

Theorem

ε-Guarded ERM using $m \gtrsim \frac{\log \frac{1}{\varepsilon\delta}}{\varepsilon^3}$ samples is an $1-\varepsilon$ approximation w.p. $1-\delta$.

- To get $\left|\frac{1}{m}\sum_{i=1}^m X_i - \mu\right| \leq \varepsilon\mu$ we need $m \gtrsim \frac{\log \frac{1}{\delta}}{\mu\varepsilon^2}$ samples
- It suffices consider prices with quantiles at least ε

Lower Bound for Regular Distributions

Theorem

Any $1-\varepsilon$ approximation algorithm uses at least $m \gtrsim \frac{1}{\varepsilon^3}$ samples.

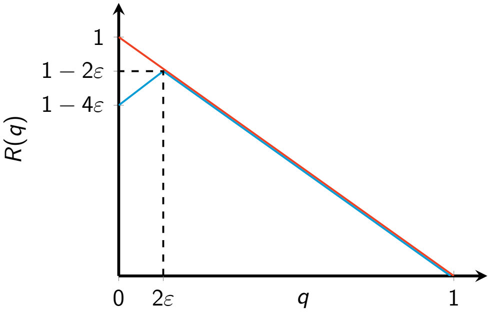
- Construct two regular value distributions P and Q that are
  1. "Similar": For ε fraciton of v, $e^{-\varepsilon} \leq \frac{P(v)}{Q(v)} \leq e^{\varepsilon}$; for the rest, $P(v) = Q(v)$
  2. "Different": No price p is a $1-\varepsilon$ approximation for both P and Q

Lower Bound for Regular Distributions

Theorem

Any $1 - \varepsilon$ approximation algorithm uses at least $m \gtrsim \frac{1}{\varepsilon^3}$ samples.

- Construct two regular value distributions P and Q that are
  1. "Similar": For ε fraciton of v, $e^{-\varepsilon} \leq \frac{P(v)}{Q(v)} \leq e^{\varepsilon}$; for the rest, $P(v) = Q(v)$
  2. "Different": No price p is a $1 - \varepsilon$ approximation for both P and Q



$[1, H]$-Bounded Distributions

Theorem
$\frac{1}{H}$*-Guarded ERM using* $m \gtrsim \frac{H \log \frac{1}{\varepsilon\delta}}{\varepsilon^2}$ *samples is an* $1 - \varepsilon$ *approximation w.p.* $1 - \delta$.

Theorem
Any $1 - \varepsilon$ *approximation algorithm uses at least* $m \gtrsim \frac{H}{\varepsilon^2}$ *samples.*

MHR Distributions

Theorem

ERM using $m \gtrsim \frac{\log \frac{1}{\varepsilon\delta}}{\varepsilon^{1.5}}$ *samples is an* $1-\varepsilon$ *approximation w.p.* $1-\delta$.

Theorem

Any $1-\varepsilon$ *approximation algorithm uses at least* $m \gtrsim \frac{1}{\varepsilon^{1.5}}$ *samples.*

Summary

Distributions	Sample Complexity
$[0, 1]$-Bounded	$\frac{1}{\varepsilon^2}$
Regular distributions	$\frac{1}{\varepsilon^3}$
MHR distributions	$\frac{1}{\varepsilon^{1.5}}$
$[1, H]$-bounded distributions	$\frac{H}{\varepsilon^2}$

- **Upper Bound:**
  Concentration inequality + covering of price space + union bound
- **Lower Bound:**
  Reduction to sample complexity of distinguishing two distributions

Summary

Distributions	Sample Complexity
$[0, 1]$-Bounded	$\frac{1}{\varepsilon^2}$
Regular distributions	$\frac{1}{\varepsilon^3}$
MHR distributions	$\frac{1}{\varepsilon^{1.5}}$
$[1, H]$-bounded distributions	$\frac{H}{\varepsilon^2}$

- **Upper Bound:**
  Concentration inequality + covering of price space + union bound
- **Lower Bound:**
  Reduction to sample complexity of distinguishing two distributions

Take-Home Question: Can we get all upper bounds using the same algorithm?

References

1. Richard Cole and Tim Roughgarden. *"The sample complexity of revenue maximization."* In Proceedings of the 46th Annual ACM Symposium on Theory of Computing, ACM, pp. 243–252, 2014.
2. Peerapong Dhangwatnotai, Tim Roughgarden, Qiqi Yan. *"Revenue maximization with a single sample."* Games and Economic Behavior 91, pp. 318–333, 2015.
3. Zhiyi Huang, Yishay Mansour, and Tim Roughgarden. *"Making the most of your samples."* SIAM Journal on Computing, 47(3), pp. 651–674, 2018.