Data-Driven Auction Design II

Progress via Statistical Learning Theory

Zhiyi Huang

University of Hong Kong

Recap

Data-Driven Single-Item Auction

A Glimpse of Statistical Learning Theory

Sample Complexity of Single-Item Auctions
- Upper Bound
- Lower Bound

Recap: Single-Item Auctions

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

Recap: Single-Item Auctions

- 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 $v_i x_i - p_i$

Recap: Single-Item Auctions

- 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)

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

Recap: Single-Item Auctions

- 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)

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

Recap: 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)

Recap: 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

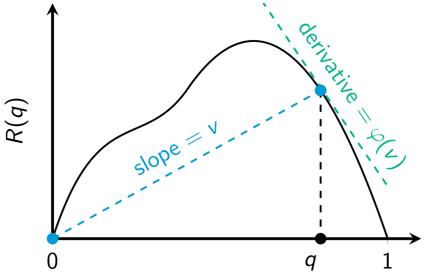
$$\mathbf{E} \sum_{i=1}^{n} \varphi_i(v_i) x_i$$

where the virtual value φ_i is

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

Recap: Optimal Pricing

- 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 $R(q) = v(q) \cdot q$



Recap: Data-Driven Optimal Pricing

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

Recap: Data-Driven Optimal Pricing

- 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
  - $1 - \varepsilon$ (multiplicative) approximation

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

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

    $[1, H]$-bounded distributions

Recap: Data-Driven Optimal Pricing

- 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
  - $1 - \varepsilon$ (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

Recap: Data-Driven Optimal Pricing (Cont'd)

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

Recap: Concentration Inequalities

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

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

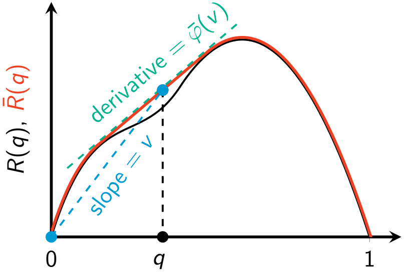
Theorem (Bernstein Inequality, User-Friendly Version)

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

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

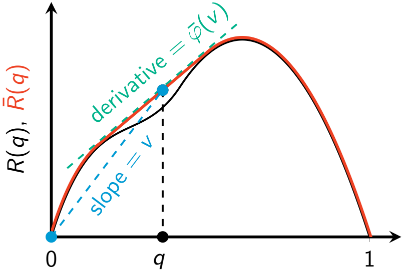
Myerson's Optimal (Single-Item) Auction

- $\bar{R}(q)$ is concave closure of revenue curve
- Ironed virtual value $\bar{\varphi}_i(v_i)$ is $\bar{R}(q)$'s derivative



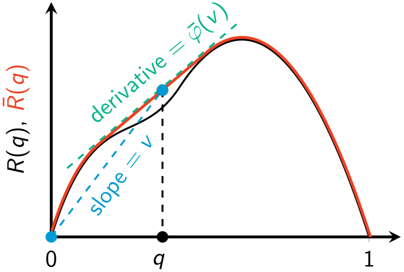
Myerson's Optimal (Single-Item) Auction

- $\bar{R}(q)$ is concave closure of revenue curve
  - Max expected revenue given sale prob. q
- Ironed virtual value $\bar{\varphi}_i(v_i)$ is $\bar{R}(q)$'s derivative



Myerson's Optimal (Single-Item) Auction

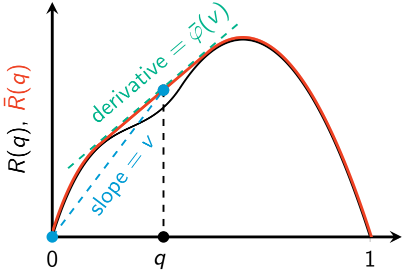
- $\bar{R}(q)$ is concave closure of revenue curve
  - Max expected revenue given sale prob. q
- Ironed virtual value $\bar{\varphi}_i(v_i)$ is $\bar{R}(q)$'s derivative
  - Quantile q's marginal revenue contribution



Myerson's Optimal (Single-Item) Auction

- $\bar{R}(q)$ is concave closure of revenue curve
  - Max expected revenue given sale prob. q
- Ironed virtual value $\bar{\varphi}_i(v_i)$ is $\bar{R}(q)$'s derivative
  - Quantile q's marginal revenue contribution

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

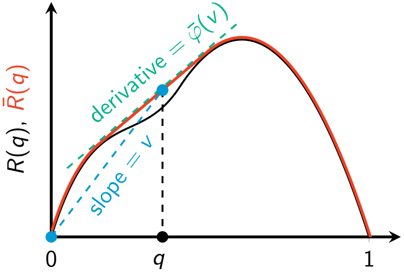


Myerson's Optimal (Single-Item) Auction

- $\bar{R}(q)$ is concave closure of revenue curve
  - Max expected revenue given sale prob. q
- Ironed virtual value $\bar{\varphi}_i(v_i)$ is $\bar{R}(q)$'s derivative
  - Quantile q's marginal revenue contribution

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

- Expected revenue is at most $\mathbf{E} \sum_{i=1}^{n} \bar{\varphi}_i(v_i) x_i$
  with equality if values in an ironed interval are treated as the same



Data-Driven Optimal (Single-Item) Auction

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from $D = D_1 \times D_2 \times \cdots \times D_m$ as input
  - Output a DSIC and IR auction A

Data-Driven Optimal (Single-Item) Auction

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from $D = D_1 \times D_2 \times \cdots \times D_m$ as input
  - Output a DSIC and IR auction A
- How many samples are needed to pick a near optimal A "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 (Single-Item) Auction

- Sample Complexity/Statistical Learning Model
  - Take m i.i.d. samples from $D = D_1 \times D_2 \times \cdots \times D_m$ as input
  - Output a DSIC and IR auction A
- How many samples are needed to pick a near optimal A "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

General Learning Problem

The learning process is a process of choosing an appropriate function from a given set of functions. —*Vladimir Vapnik*

General Learning Problem

The learning process is a process of choosing an appropriate function from a given set of functions. —*Vladimir Vapnik*

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$

General Learning Problem

> *The learning process is a process of choosing an appropriate function from a given set of functions.* —*Vladimir Vapnik*

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

General Learning Problem

> *The learning process is a process of choosing an appropriate function from a given set of functions.* —*Vladimir Vapnik*

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$
- Learn $h \in \mathcal{H}$ from i.i.d. samples from D to minimize/maximize

$$\mathbf{E}_{t \sim D}\, h(t)$$

Example: Linear Binary Classification

- Type space consists of feature-label pairs

$$\mathcal{T} = \{\ (x,y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}}\ \}$$

Example: Linear Binary Classification

- Type space consists of feature-label pairs

$$\mathcal{T} = \{\ (x, y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}}\ \}$$

- Hypothesis space consists of linear classifiers
  - Each $h \in \mathcal{H}$ corresponds to a linear function $\langle a, x \rangle + b$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$h(x, y) = \begin{cases} 0 & \text{if } \langle a, x \rangle + b \text{ and } y \text{ have the same sign} \\ 1 & \text{otherwise} \end{cases}$$

Example: Linear Binary Classification

- Type space consists of feature-label pairs

$$\mathcal{T} = \{\ (x, y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}}\ \}$$

- Hypothesis space consists of linear classifiers
  - Each $h \in \mathcal{H}$ corresponds to a linear function $\langle a, x \rangle + b$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

$$h(x, y) = \begin{cases} 0 & \text{if } \langle a, x \rangle + b \text{ and } y \text{ have the same sign} \\ 1 & \text{otherwise} \end{cases}$$

- Learn $h \in \mathcal{H}$ from i.i.d. samples from D to minimize $\underbrace{\mathbf{E}_{(x,y)\sim D}\, h(x, y)}_{\text{classfication error}}$

Example: Data-Driven Optimal Auction

- Type space consists of value profiles of n bidders

$$\mathcal{T} = [0, 1]^n$$

Example: Data-Driven Optimal Auction

- Type space consists of value profiles of n bidders

$$\mathcal{T} = [0, 1]^n$$

- Hypothesis space consists of DSIC and IR auctions
  - Each $h \in \mathcal{H}$ corresponds to a DSIC and IR auciton A
  - $h(v)$ equals the expected revenue of running A on values v

Example: Data-Driven Optimal Auction

- Type space consists of value profiles of n bidders

$$\mathcal{T} = [0, 1]^n$$

- Hypothesis space consists of DSIC and IR auctions
  - Each $h \in \mathcal{H}$ corresponds to a DSIC and IR auciton A
  - $h(v)$ equals the expected revenue of running A on values v

- Learn $h \in \mathcal{H}$ from i.i.d. samples from D to maximize $\underbrace{\mathbf{E}_{v \sim D}\, h(v)}_{\text{expected revenue}}$

Sample Complexity and "Degree of Freedom": Informal Introduction

Recall the three-step approach

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

3. Estimate the expectations of all these representative hypotheses up to ε

Sample Complexity and "Degree of Freedom": Informal Introduction

Recall the three-step approach

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   Solution: Chernoff-Hoeffding Bound, Bernstein Inequality

   Takeaway: $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

3. Estimate the expectations of all these representative hypotheses up to ε

Sample Complexity and "Degree of Freedom": Informal Introduction

Recall the three-step approach

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   Solution: Chernoff-Hoeffding Bound, Bernstein Inequality

   Takeaway: $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

3. Estimate the expectations of all these representative hypotheses up to ε

   Solution: Union Bound

   Takeaway: $m \gtrsim \frac{\log \frac{R}{\delta}}{\varepsilon^2}$ samples suffice when there are R representative hypotheses

Sample Complexity and "Degree of Freedom": Informal Introduction

Recall the three-step approach

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   Solution: Chernoff-Hoeffding Bound, Bernstein Inequality

   Takeaway: $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

   Conventional wisdom: If the hypothesis space $\mathcal{H}$ has "degree of freedom" d (a.k.a., "dimension"), then $R = 2^{\tilde{O}(d)}$ representative hypotheses suffice

3. Estimate the expectations of all these representative hypotheses up to ε

   Solution: Union Bound

   Takeaway: $m \gtrsim \frac{\log \frac{R}{\delta}}{\varepsilon^2}$ samples suffice when there are R representative hypotheses

Sample Complexity and "Degree of Freedom": Informal Introduction

Recall the three-step approach

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   Solution: Chernoff-Hoeffding Bound, Bernstein Inequality

   Takeaway: $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

   Conventional wisdom: If the hypothesis space $\mathcal{H}$ has "degree of freedom" d (a.k.a., "dimension"), then $R = 2^{\tilde{O}(d)}$ representative hypotheses suffice

3. Estimate the expectations of all these representative hypotheses up to ε

   Solution: Union Bound

   Takeaway: $m \gtrsim \frac{\log \frac{R}{\delta}}{\varepsilon^2}$ samples suffice when there are R representative hypotheses

   i.e., $m \gtrsim \frac{d \log \frac{1}{\delta}}{\varepsilon^2}$ samples suffice

Binary Classification and Vapnik-Chervonenkis Dimension

- Type space consists of feature-label pairs

$$\mathcal{T} = \{ \ (x, y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}} \ \}$$

- Hypothesis space is a set of classifiers
  - Each $h \in \mathcal{H}$ corresponds to a classifier $c : \mathbb{R}^n \to \{-1, +1\}$

$$h(x, y) = \begin{cases} 0 & \text{if } c(x) = y \\ 1 & \text{otherwise} \end{cases}$$

Binary Classification and Vapnik-Chervonenkis Dimension

- Type space consists of feature-label pairs

$$\mathcal{T} = \{ \ (x, y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}} \ \}$$

- Hypothesis space is a set of classifiers
  - Each $h \in \mathcal{H}$ corresponds to a classifier $c : \mathbb{R}^n \to \{-1, +1\}$

$$h(x, y) = \begin{cases} 0 & \text{if } c(x) = y \\ 1 & \text{otherwise} \end{cases}$$

- VC dimension of $\mathcal{H}$ is the largest number of features vectors $x_1, x_2, \ldots, x_d$ such that for any labeling $y_1, y_2, \ldots, y_d$, there is $h \in H$ such that $h(x_i, y_i) = 0$

Binary Classification and Vapnik-Chervonenkis Dimension

- Type space consists of feature-label pairs

$$\mathcal{T} = \{\ (x, y) : \underbrace{x \in \mathbb{R}^n}_{\text{feature}}, \underbrace{y = \pm 1}_{\text{label}}\ \}$$

- Hypothesis space is a set of classifiers
  - Each $h \in \mathcal{H}$ corresponds to a classifier $c : \mathbb{R}^n \to \{-1, +1\}$

$$h(x, y) = \begin{cases} 0 & \text{if } c(x) = y \\ 1 & \text{otherwise} \end{cases}$$

- VC dimension of $\mathcal{H}$ is the largest number of features vectors $x_1, x_2, \ldots, x_d$ such that for any labeling $y_1, y_2, \ldots, y_d$, there is $h \in H$ such that $h(x_i, y_i) = 0$
- $m \gtrsim \frac{d + \log \frac{1}{\delta}}{\varepsilon^2}$ samples are sufficient and necessary

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Pseudo dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$\text{sign}\big(h(t_i) - r_i\big) = y_i$$

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Pseudo dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$\text{sign}\big(h(t_i) - r_i\big) = y_i$$

- $m \gtrsim \frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2}$ samples are sufficient

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Pseudo dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$\text{sign}\big(h(t_i) - r_i\big) = y_i$$

- $m \gtrsim \frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2}$ samples are sufficient
- Pseudo dimension equals VC dimension for binary classification

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- (ε-)Fat shattering dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$y_i \big(h(t_i) - r_i \big) \geq \varepsilon$$

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- (ε-)Fat shattering dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$y_i \left(h(t_i) - r_i\right) \geq \varepsilon$$

- $m \gtrsim \frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2}$ samples are sufficient

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- (ε-)Fat shattering dimension of $\mathcal{H}$ is the largest number d for which we have
  - Types $t_1, t_2, \ldots, t_d \in \mathcal{T}$ and
  - Witnesses $r_1, r_2, \ldots, r_d \in (0, 1)$ such that
  - For any signs $y_1, y_2, \ldots, y_d \in \{1, -1\}$ there is $h \in \mathcal{H}$ satisfying

$$y_i\big(h(t_i) - r_i\big) \geq \varepsilon$$

- $m \gtrsim \frac{d \log \frac{1}{\varepsilon} + \log \frac{1}{\delta}}{\varepsilon^2}$ samples are sufficient
- Fat shattering dimension equals VC dimension for binary classification

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Rademacher complexity of $\mathcal{H}$ (with m samples) is

$$R_m(\mathcal{H}) = \mathbf{E}_{\underbrace{t_1,\ldots,t_m \sim D}_{\text{random types}},\ \underbrace{y_1,y_2,\ldots,y_m \overset{unif}{\sim} \{1,-1\}}_{\text{random noise}}} \sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} y_i h(t_i)$$

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Rademacher complexity of $\mathcal{H}$ (with m samples) is

$$R_m(\mathcal{H}) = \mathbf{E}_{\underbrace{t_1,\ldots,t_m \sim D}_{\text{random types}}, \underbrace{y_1,y_2,\ldots,y_m \overset{unif}{\sim} \{1,-1\}}_{\text{random noise}}} \sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} y_i h(t_i)$$

- Intuitively, it captures how well hypothesis class $\mathcal{H}$ can fit random noise

"Degree of Freedom" for General Learning Problem

- Type space $\mathcal{T}$
  - Distribution D over $\mathcal{T}$
- Hypothesis space $\mathcal{H}$
  - Each hypothesis $h \in \mathcal{H}$ is a function from $\mathcal{T}$ to $[0, 1]$

- Rademacher complexity of $\mathcal{H}$ (with m samples) is

$$R_m(\mathcal{H}) = \mathbf{E}_{\underbrace{t_1,\ldots,t_m \sim D}_{\text{random types}}, \underbrace{y_1,y_2,\ldots,y_m \overset{unif}{\sim} \{1,-1\}}_{\text{random noise}}} \sup_{h \in \mathcal{H}} \frac{1}{m} \sum_{i=1}^{m} y_i h(t_i)$$

- Intuitively, it captures how well hypothesis class $\mathcal{H}$ can fit random noise
- It suffices to have $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ and $R_m(\mathcal{H}) \lesssim \varepsilon$

Recap

Data-Driven Single-Item Auction

A Glimpse of Statistical Learning Theory

Sample Complexity of Single-Item Auctions

- Upper Bound
- Lower Bound

Explicit Covering for Single-Item Auction: $[0, 1]$-Bounded Case

Recall Myerson's optimal auction

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

Narrowing down the representitive auctions in three steps

Explicit Covering for Single-Item Auction: $[0,1]$-Bounded Case

Recall Myerson's optimal auction

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

Narrowing down the representitive auctions in three steps

1. Focus on auctions defined by n non-decreasing "ironed virtual value functions"

$$f_i : [0,1] \to (-\infty, 1]$$

Explicit Covering for Single-Item Auction: $[0,1]$-Bounded Case

Recall Myerson's optimal auction

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

Narrowing down the representitive auctions in three steps

1. Focus on auctions defined by n non-decreasing "ironed virtual value functions"
2. Discretize the value space

$$f_i : \{0, \varepsilon, 2\varepsilon, \dots, 1\} \to (-\infty, 1]$$

Explicit Covering for Single-Item Auction: $[0, 1]$-Bounded Case

Recall Myerson's optimal auction

- Highest non-negative virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins

Narrowing down the representitive auctions in three steps

1. Focus on auctions defined by n non-decreasing "ironed virtual value functions"
2. Discretize the value space
3. Discretize the virtual value space

$$f_i : \{0, \varepsilon, 2\varepsilon, \ldots, 1\} \to \{-\infty, 0, \varepsilon, 2\varepsilon, \ldots, 1\}$$

Discretization of Value Space

Lemma

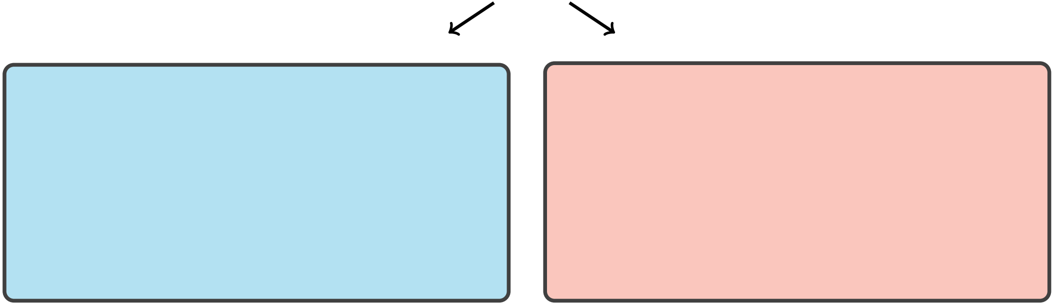
There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

Discretization of Value Space

Lemma

There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

quantiles $q_1, q_2, \ldots, q_n$



Discretization of Value Space

Lemma

There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

quantiles $q_1, q_2, \ldots, q_n$

1. Values $v_1, v_2, \ldots, v_n$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_{\boldsymbol{i}}(\boldsymbol{v_i})$
3. Winner pays threshold bid

Discretization of Value Space

Lemma

There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

quantiles $q_1, q_2, \ldots, q_n$

↙ ↘

1. Values $v_1, v_2, \ldots, v_n$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_{\boldsymbol{i}}(\boldsymbol{v_i})$
3. Winner pays threshold bid

1. Values $\lfloor v_1 \rfloor_\varepsilon, \lfloor v_2 \rfloor_\varepsilon, \ldots, \lfloor v_n \rfloor_\varepsilon$

Discretization of Value Space

Lemma

There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

quantiles $q_1, q_2, \ldots, q_n$

↙ ↘

1. Values $v_1, v_2, \ldots, v_n$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_{\boldsymbol{i}}(\boldsymbol{v_i})$
3. Winner pays threshold bid

1. Values $\lfloor v_1 \rfloor_\varepsilon, \lfloor v_2 \rfloor_\varepsilon, \ldots, \lfloor v_n \rfloor_\varepsilon$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_{\boldsymbol{i}}(\boldsymbol{v_i})$

Discretization of Value Space

Lemma

There is an auction A on value space $\{0, \varepsilon, 2\varepsilon, \ldots, 1\}$ such that rounding each v_i to the closest multiple of ε from below, denoted as $\lfloor v_i \rfloor_\varepsilon$, and running A is optimal up to ε.

quantiles $q_1, q_2, \ldots, q_n$

↙ ↘

1. Values $v_1, v_2, \ldots, v_n$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_i(v_i)$
3. Winner pays threshold bid

1. Values $\lfloor v_1 \rfloor_\varepsilon, \lfloor v_2 \rfloor_\varepsilon, \ldots, \lfloor v_n \rfloor_\varepsilon$
2. Allocate to bidder i with highest non-negative $\bar{\varphi}_i(v_i)$
3. Winner pays threshold bid, at worst smaller by ε

Discretization of Virtual Value Space

Lemma

1. *Negative $f_i(v_i)$ may be treated as $-\infty$ without loss of generality.*
2. *Allocating to the bidder i with the largest $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ breaking ties, say, lexicographically, (and letting it pay threshold bid) is optimal up to ε.*

Discretization of Virtual Value Space

Lemma

1. *Negative* $f_i(v_i)$ *may be treated as* $-\infty$ *without loss of generality.*
2. *Allocating to the bidder* i *with the largest* $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ *breaking ties, say, lexicographically, (and letting it pay threshold bid) is optimal up to* ε.

Recall: Expected revenue is

$$\mathbf{E} \sum_{i=1}^{n} \bar{\varphi}_i(v_i) x_i$$

if values in an ironed interval are treated as the same

Discretization of Virtual Value Space

Lemma

1. *Negative $f_i(v_i)$ may be treated as $-\infty$ without loss of generality.*
2. *Allocating to the bidder i with the largest $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ breaking ties, say, lexicographically, (and letting it pay threshold bid) is optimal up to ε.*

Recall: Expected revenue is

$$\mathbf{E} \sum_{i=1}^{n} \bar{\varphi}_i(v_i) x_i$$

if values in an ironed interval are treated as the same

- Allocating to largest $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ still treats values in an ironed interval as the same

Discretization of Virtual Value Space

Lemma

1. *Negative $f_i(v_i)$ may be treated as $-\infty$ without loss of generality.*
2. *Allocating to the bidder i with the largest $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ breaking ties, say, lexicographically, (and letting it pay threshold bid) is optimal up to ε.*

Recall: Expected revenue is

$$\mathbf{E} \sum_{i=1}^{n} \bar{\varphi}_i(v_i) x_i$$

if values in an ironed interval are treated as the same

- Allocating to largest $\lfloor \bar{\varphi}_i(v_i) \rfloor_\varepsilon$ still treats values in an ironed interval as the same
- Lose at most ε in $\mathbf{E} \sum_{i=1}^{n} \bar{\varphi}_i(v_i) x_i$

Information Theoretic Upper Bound

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, we can find an auction that is an* ε *additive approximation with probability* $1 - \delta$.

Information Theoretic Upper Bound

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, we can find an auction that is an* ε *additive approximation with probability* $1 - \delta$.

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

3. Estimate the expectations of all these representative hypotheses up to ε

   $m \gtrsim \frac{\log \frac{R}{\delta}}{\varepsilon^2}$ samples suffice when there are R representative hypotheses

Information Theoretic Upper Bound

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, we can find an auction that is an* ε *additive approximation with probability* $1 - \delta$*.*

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

2. Finitely many hypotheses whose "neighborhoods" cover the hypothesis space $\mathcal{H}$

   Focus on $R \lesssim (\frac{1}{\varepsilon} + 2)^{n(\frac{1}{\varepsilon}+1)}$ auctions defined by n non-decreasing functions

$$f_i : \{0, \varepsilon, 2\varepsilon, \ldots, 1\} \to \{-\infty, 0, \varepsilon, 2\varepsilon, \ldots, 1\}$$

3. Estimate the expectations of all these representative hypotheses up to ε

   $m \gtrsim \frac{\log \frac{R}{\delta}}{\varepsilon^2}$ samples suffice when there are R representative hypotheses

Information Theoretic Upper Bound

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, we can find an auction that is an* ε *additive approximation with probability* $1 - \delta$*.*

1. Estimate the expectation of a single hypothesis $h \in \mathcal{H}$ up to ε

   $m \gtrsim \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples give ε additive approximation w.p. $1 - \delta$

Why information theoretic?

- We estimate revenue by averaging over samples, i.e., empirical distribution
- Empirical distribution is not independent
- Optimal auction over dependent value distribution is hard

Upper Bound via Polynomial-Time Algorithm

Empirical Myerson's Auction (with Value Discretization)

- Given i.i.d. samples $v^i = (v_1^i, v_2^i, \ldots, v_n^i)$, $1 \leq i \leq m$
- Let E_j be the uniform distribution over $\lfloor v_j^1 \rfloor_\varepsilon, \lfloor v_j^2 \rfloor_\varepsilon, \ldots, \lfloor v_j^m \rfloor_\varepsilon$
- Return Myerson's optimal auction w.r.t. $E = E_1 \times E_2 \times \cdots \times E_n$

Upper Bound via Polynomial-Time Algorithm

Empirical Myerson's Auction (with Value Discretization)

- Given i.i.d. samples $v^i = (v_1^i, v_2^i, \ldots, v_n^i)$, $1 \le i \le m$
- Let E_j be the uniform distribution over $\lfloor v_j^1 \rfloor_\varepsilon, \lfloor v_j^2 \rfloor_\varepsilon, \ldots, \lfloor v_j^m \rfloor_\varepsilon$
- Return Myerson's optimal auction w.r.t. $E = E_1 \times E_2 \times \cdots \times E_n$

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, Empirical Myerson's Auction is an* ε *additive approximation with probability* $1 - \delta$.

Upper Bound via Polynomial-Time Algorithm

Empirical Myerson's Auction (with Value Discretization)

- Given i.i.d. samples $v^i = (v_1^i, v_2^i, \ldots, v_n^i)$, $1 \le i \le m$
- Let E_j be the uniform distribution over $\lfloor v_j^1 \rfloor_\varepsilon, \lfloor v_j^2 \rfloor_\varepsilon, \ldots, \lfloor v_j^m \rfloor_\varepsilon$
- Return Myerson's optimal auction w.r.t. $E = E_1 \times E_2 \times \cdots \times E_n$

Theorem

Using $m \gtrsim \frac{n \log \frac{1}{\varepsilon}}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2} = \tilde{O}(\frac{n}{\varepsilon^3})$ *samples, Empirical Myerson's Auction is an* ε *additive approximation with probability* $1 - \delta$.

Lemma (Bernstein Inequality for Product Distribution)

For any function $f : [0,1]^n \to [0,1]$. *Let* $\mu = \mathbf{E}_{v \sim E} f(v)$. *With probability* $1 - \delta$

$$\left| \mathbf{E}_{v \sim E} f(v) - \mu \right| \lesssim \max \left\{ \sqrt{\frac{\mu(1-\mu) \log \frac{1}{\delta}}{m}}, \frac{\log \frac{1}{\delta}}{m} \right\}$$

Lower Bound

Theorem

Any ε additive approximation algorithm needs at least $m \gtrsim \frac{n}{\varepsilon^2}$ *samples.*

Lower Bound

Theorem

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

- Linear dependence in number of bidders n
  - **Upshot:** The multi-bidder problem is strictly harder

Lower Bound

Theorem

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

- Linear dependence in number of bidders n
  - **Upshot:** The multi-bidder problem is strictly harder
  - Note that we already let each sample be a vector of n values

Lower Bound

Theorem

Any ε additive approximation algorithm needs at least $m \gtrsim \frac{n}{\varepsilon^2}$ *samples.*

- Linear dependence in number of bidders n
  - **Upshot:** The multi-bidder problem is strictly harder
  - Note that we already let each sample be a vector of n values
  - We need more information about each bidder's value distribtuion

Lower Bound

Theorem

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

- Linear dependence in number of bidders n
  - **Upshot:** The multi-bidder problem is strictly harder
  - Note that we already let each sample be a vector of n values
  - We need more information about each bidder's value distribtuion
- Dependence on ε does not match the upper bound, i.e., quadratic vs. cubic
  - Next lecture will resolve this gap

Le Cam's Method is Insufficient

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

Le Cam's Method is Insufficient

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

- **1st Attempt**
  - Make P, Q similar s.t. distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Any auction's revenue differ by $\lesssim \frac{\varepsilon}{\sqrt{n}}$ on P and Q

Le Cam's Method is Insufficient

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

- **1st Attempt**
  - Make P, Q similar s.t. distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Any auction's revenue differ by $\lesssim \frac{\varepsilon}{\sqrt{n}}$ on P and Q

- **2nd Attempt**
  - Make marginals P_i, Q_i similar s.t. distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - It takes much fewer samples to distinguish product distributions P and Q

Assouad's Method

- Two marginals P, Q for each bidder, distinguishing which needs $m \gtrsim \frac{n}{\varepsilon^2}$ samples

Assouad's Method

- Two marginals P, Q for each bidder, distinguishing which needs $m \gtrsim \frac{n}{\varepsilon^2}$ samples
- Consider 2^n value distributions $D = D_1 \times D_2 \times \cdots \times D_n$, where $D_i \in \{P, Q\}$

Assouad's Method

- Two marginals P, Q for each bidder, distinguishing which needs $m \gtrsim \frac{n}{\varepsilon^2}$ samples
- Consider 2^n value distributions $D = D_1 \times D_2 \times \cdots \times D_n$, where $D_i \in \{P, Q\}$
- Argument sketch
  - For neighboring D, D' differing in bidder i's marginal, any algorithm "makes some mistake" in i's allocation, resulting in $\gtrsim \frac{\varepsilon}{n}$ total revenue loss to D, D'

Assouad's Method

- Two marginals P, Q for each bidder, distinguishing which needs $m \gtrsim \frac{n}{\varepsilon^2}$ samples
- Consider 2^n value distributions $D = D_1 \times D_2 \times \cdots \times D_n$, where $D_i \in \{P, Q\}$
- Argument sketch
  - For neighboring D, D' differing in bidder i's marginal, any algorithm "makes some mistake" in i's allocation, resulting in $\gtrsim \frac{\varepsilon}{n}$ total revenue loss to D, D'
  - $2^{n-1}n$ pairs of neighboring distributions

Assouad's Method

- Two marginals P, Q for each bidder, distinguishing which needs $m \gtrsim \frac{n}{\varepsilon^2}$ samples
- Consider 2^n value distributions $D = D_1 \times D_2 \times \cdots \times D_n$, where $D_i \in \{P, Q\}$
- Argument sketch
  - For neighboring D, D' differing in bidder i's marginal, any algorithm "makes some mistake" in i's allocation, resulting in $\gtrsim \frac{\varepsilon}{n}$ total revenue loss to D, D'
  - $2^{n-1}n$ pairs of neighboring distributions
  - Some distribution D has revenue loss at least

$$\underbrace{2^{n-1}n}_{\text{number of neighboring pairs}} \cdot \underbrace{\frac{\varepsilon}{n}}_{\text{revenue loss per pair}} \cdot \underbrace{\frac{1}{2^n}}_{\text{divided by number of distributions}} \eqsim \varepsilon$$

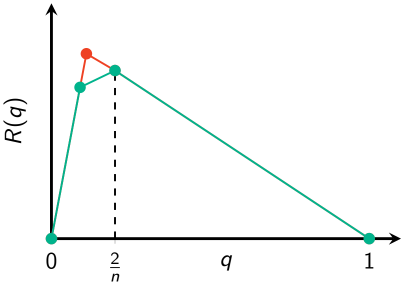
Assouad's Method (cont'd)

- P and Q have support $\left\{0, \frac{1}{2}, 1\right\}$

Assouad's Method (cont'd)

- P and Q have support $\{0, \frac{1}{2}, 1\}$

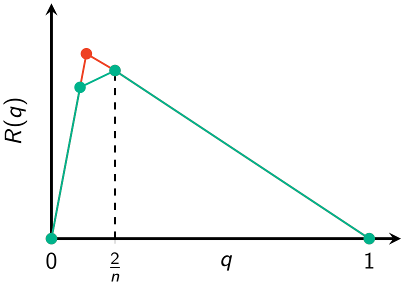
v	1	$\frac{1}{2}$	0
$P(v)$	$\frac{1+\varepsilon}{n}$	$\frac{1-\varepsilon}{n}$	$1-\frac{2}{n}$
$Q(v)$	$\frac{1-\varepsilon}{n}$	$\frac{1+\varepsilon}{n}$	$1-\frac{2}{n}$



Assouad's Method (cont'd)

- P and Q have support $\{0, \frac{1}{2}, 1\}$
- Distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Differ only in $\frac{2}{n}$ of the mass
  - Differ by at most $1 \pm \varepsilon$ for any v
  - $\mathrm{KL}(P, Q) \lesssim \frac{\varepsilon^2}{n}$ (last lecture)

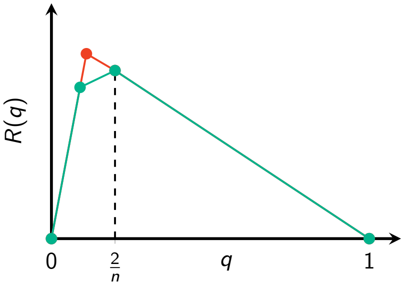
v	1	$\frac{1}{2}$	0
P(v)	$\frac{1+\varepsilon}{n}$	$\frac{1-\varepsilon}{n}$	$1 - \frac{2}{n}$
Q(v)	$\frac{1-\varepsilon}{n}$	$\frac{1+\varepsilon}{n}$	$1 - \frac{2}{n}$



Assouad's Method (cont'd)

- P and Q have support $\left\{0, \frac{1}{2}, 1\right\}$
- Distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Differ only in $\frac{2}{n}$ of the mass
  - Differ by at most $1 \pm \varepsilon$ for any v
  - $\mathrm{KL}(P, Q) \lesssim \frac{\varepsilon^2}{n}$ (last lecture)
- Revenue loss due to $D_i = P$ vs. $D_i = Q$
  - w.p. $\approx \frac{1}{n}$, $v_i = \frac{1}{2}$ and other values are zero

v	1	$\frac{1}{2}$	0
P(v)	$\frac{1+\varepsilon}{n}$	$\frac{1-\varepsilon}{n}$	$1 - \frac{2}{n}$
Q(v)	$\frac{1-\varepsilon}{n}$	$\frac{1+\varepsilon}{n}$	$1 - \frac{2}{n}$

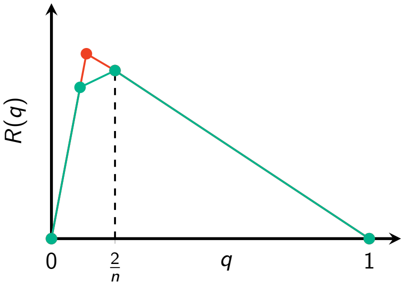


Assoud's Method (cont'd)

- P and Q have support $\{0, \frac{1}{2}, 1\}$
- Distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Differ only in $\frac{2}{n}$ of the mass
  - Differ by at most $1 \pm \varepsilon$ for any v
  - $\mathrm{KL}(P, Q) \lesssim \frac{\varepsilon^2}{n}$ (last lecture)
- Revenue loss due to $D_i = P$ vs. $D_i = Q$
  - w.p. $\approx \frac{1}{n}$, $v_i = \frac{1}{2}$ and other values are zero
  - Bidder i should win iff. $D_i = Q$

$$\varphi_i\left(\tfrac{1}{2}\right) \approx \begin{cases} -\varepsilon & D_i = P \\ \varepsilon & D_i = Q \end{cases}$$

v	1	$\frac{1}{2}$	0
P(v)	$\frac{1+\varepsilon}{n}$	$\frac{1-\varepsilon}{n}$	$1 - \frac{2}{n}$
Q(v)	$\frac{1-\varepsilon}{n}$	$\frac{1+\varepsilon}{n}$	$1 - \frac{2}{n}$



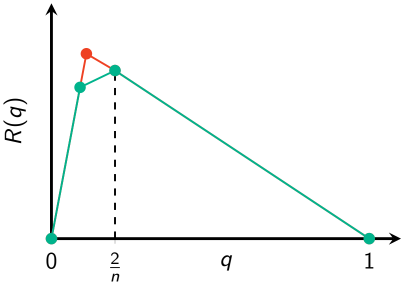
Assouad's Method (cont'd)

- P and Q have support $\{0, \frac{1}{2}, 1\}$
- Distinguishing them takes $m \gtrsim \frac{n}{\varepsilon^2}$ samples
  - Differ only in $\frac{2}{n}$ of the mass
  - Differ by at most $1 \pm \varepsilon$ for any v
  - $\mathrm{KL}(P, Q) \lesssim \frac{\varepsilon^2}{n}$ (last lecture)
- Revenue loss due to $D_i = P$ vs. $D_i = Q$
  - w.p. $\approx \frac{1}{n}$, $v_i = \frac{1}{2}$ and other values are zero
  - Bidder i should win iff. $D_i = Q$

$$\varphi_i\left(\tfrac{1}{2}\right) \approx \begin{cases} -\varepsilon & D_i = P \\ \varepsilon & D_i = Q \end{cases}$$

  - Lose $\gtrsim \frac{\varepsilon}{n}$ if we cannot distinguish P, Q

v	1	$\frac{1}{2}$	0
P(v)	$\frac{1+\varepsilon}{n}$	$\frac{1-\varepsilon}{n}$	$1 - \frac{2}{n}$
Q(v)	$\frac{1-\varepsilon}{n}$	$\frac{1+\varepsilon}{n}$	$1 - \frac{2}{n}$



Summary

Distributions	Upper Bound	Lower Bound
$[0,1]$-Bounded	$\frac{n}{\varepsilon^3}$	$\frac{n}{\varepsilon^2}$
Regular distributions	$\frac{n}{\varepsilon^4}$	$\frac{n}{\varepsilon^3}$
MHR distributions	$\frac{n}{\varepsilon^3}$	$\frac{n}{\varepsilon^2}$
$[1, H]$-bounded distributions	$\frac{Hn}{\varepsilon^3}$	$\frac{Hn}{\varepsilon^2}$

- **Upper Bound:**
  Concentration inequality + covering of auction space + union bound
- **Lower Bound:**
  Assouad's Method

References

1. Jamie Morgenstern and Tim Roughgarden. *"The pseudo-dimension of near-optimal auctions."* In Proceedings of the 28th International Conference on Neural Information Processing Systems, pp. 136–144, 2015.
2. Nikhil R Devanur, Zhiyi Huang, and Christos-Alexandros Psomas. *"The sample complexity of auctions with side information."* In Proceedings of the 48th Annual ACM Symposium on Theory of Computing, ACM, pp. 426–439, 2016.
3. Yannai A Gonczarowski and Noam Nisan. *"Efficient empirical revenue maximization in single-parameter auction environments."* In Proceedings of the 49th Annual ACM Symposium on Theory of Computing, ACM, pp. 856–868, 2017.