Data-Driven Auction Design III
Learnability of Product Distributions and Strong Revenue Monotonicity

Zhiyi Huang
University of Hong Kong
Recap

Two Different Viewpoints

Learnability of Product Distributions

Strong (Revenue) Monotonicity

Further Extensions and Open Questions
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

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

where the virtual value $\varphi_i$ is

$$
\varphi_i(v_i) = v_i - \frac{1 - F_i(v_i)}{f_i(v_i)}
$$
Recap: Myerson’s Optimal (Single-Item) Auction

- $\tilde{R}(q)$ is concave closure of revenue curve
  - Max expected revenue given sale prob. $q$
- Ironed virtual value $\tilde{\varphi}_i(v_i)$ is $\tilde{R}(q)$’s derivative
  - Quantile $q$’s marginal revenue contribution
- Highest non-negative ironed virtual value wins
- Winner pays threshold winning bid
  i.e., lowest bid above which he/she wins
- Expected revenue is at most $E \sum_{i=1}^{n} \tilde{\varphi}_i(v_i)x_i$
  with equality if values in an ironed interval are treated as the same
Recap: 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 $\varepsilon$ margin”?
  - $\varepsilon$ 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
Recap: Summary of Upper and Lower Bounds So Far

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Upper Bound</th>
<th>Lower Bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, 1]-Bounded</td>
<td>( \frac{n}{\varepsilon^3} )</td>
<td>( \frac{n}{\varepsilon^2} )</td>
</tr>
<tr>
<td>Regular distributions</td>
<td>( \frac{n}{\varepsilon^4} )</td>
<td>( \frac{n}{\varepsilon^3} )</td>
</tr>
<tr>
<td>MHR distributions</td>
<td>( \frac{n}{\varepsilon^3} )</td>
<td>( \frac{n}{\varepsilon^2} )</td>
</tr>
<tr>
<td>[1, (H)]-bounded distributions</td>
<td>( \frac{Hn}{\varepsilon^3} )</td>
<td>( \frac{Hn}{\varepsilon^2} )</td>
</tr>
</tbody>
</table>

- **Upper Bound:**
  Concentration inequality + covering of auction space + union bound

- **Lower Bound:**
  Assouad’s method
Recap: Concentration Inequalities

Theorem (Chernoff-Hoeffding, User-Friendly Version)

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

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

Theorem (Bernstein Inequality, User-Friendly Version)

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

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

Two Different Viewpoints

Learnability of Product Distributions

Strong (Revenue) Monotonicity

Further Extensions and Open Questions
Recall our approach for data-driven optimal pricing
- It suffices to learn the revenue of every price up to $\varepsilon$
- For each price, estimating its revenue reduces to estimating its quantile

Next consider an alternative approach
- It suffices to learn the distribution up to $\varepsilon$ w.r.t. its CDF/quantile, i.e.
  $$\sup_{v \in [0, 1]} \left| F_D(v) - F_E(v) \right| \leq \varepsilon$$
  (Kolmogorov distance)

Then, for each price, we know its revenue up to $\varepsilon$
Recall our approach for data-driven optimal pricing
- It suffices to learn the revenue of every price up to $\varepsilon$
- For each price, estimating its revenue reduces to estimating its quantile

Next consider an alternative approach
Recall our approach for data-driven optimal pricing
- It suffices to learn the revenue of every price up to $\varepsilon$
- For each price, estimating its revenue reduces to estimating its quantile

Next consider an alternative approach
- It suffices to learn the distribution up to $\varepsilon$ w.r.t. its CDF/quantile, i.e.

$$\sup_{v \in [0,1]} \left| \hat{F}(v) - \hat{F}(v) \right| \leq \varepsilon$$  (Kolmogorov distance)
Recall our approach for data-driven optimal pricing
- It suffices to learn the revenue of every price up to $\varepsilon$
- For each price, estimating its revenue reduces to estimating its quantile

Next consider an alternative approach
- It suffices to learn the distribution up to $\varepsilon$ w.r.t. its CDF/quantile, i.e.

$$\sup_{v \in [0,1]} \left| F_D(v) - F_E(v) \right| \leq \varepsilon \quad \text{(Kolmogorov distance)}$$

- Then, for each price, we know its revenue up to $\varepsilon$
Learning Prices’ Revenue vs. Learning Value Distribution

- Recall our approach for data-driven optimal pricing
  - It suffices to learn the revenue of every price up to $\varepsilon$
  - For each price, estimating its revenue reduces to estimating its quantile

- Next consider an alternative approach
  - It suffices to learn the distribution up to $\varepsilon$ w.r.t. its CDF/quantile, i.e.

$$\sup_{v \in [0,1]} \left| F_D(v) - F_E(v) \right| \leq \varepsilon$$

(Kolmogorov distance)

- Then, for each price, we know its revenue up to $\varepsilon$

- Two approaches coincide for pricing...
Recall our approach for data-driven optimal pricing

- It suffices to learn the revenue of every price up to $\varepsilon$
- For each price, estimating its revenue reduces to estimating its quantile

Next consider an alternative approach

- It suffices to learn the distribution up to $\varepsilon$ w.r.t. its CDF/quantile, i.e.

$$\sup_{v \in [0,1]} \left| F_D(v) - F_E(v) \right| \leq \varepsilon$$

(Kolmogorov distance)

- Then, for each price, we know its revenue up to $\varepsilon$

Two approaches coincide for pricing... but not for auctions
Lower Confidence (Revenue) Bounds vs. Underestimating Distribution

- Recall the take-home question regarding optimal pricing
  - Different value distributions require different regularization in Lecture I
  - Can we get all upper bounds using the same algorithm?
Lower Confidence (Revenue) Bounds vs. Underestimating Distribution

- Recall the take-home question regarding optimal pricing
  - Different value distributions require different regularization in Lecture I
  - Can we get all upper bounds using the same algorithm?

- Lower Confidence Bound (LCB), e.g., choose $p$ to maximize

$$p \cdot \left( \frac{\left| \text{number of samples} \geq p \right|}{m} - \sqrt{\frac{\log \frac{1}{\delta}}{m}} \right)$$
Lower Confidence (Revenue) Bounds vs. Underestimating Distribution

- Recall the take-home question regarding optimal pricing
  - Different value distributions require different regularization in Lecture I
  - Can we get all upper bounds using the same algorithm?

- Lower Confidence Bound (LCB), e.g., choose $p$ to maximize
  \[
  p \cdot \left( \frac{\left| \text{number of samples} \geq p \right|}{m} - \sqrt{\log \frac{1}{\delta}} \right)
  \]

- Alternatively, consider an underestimation of the value distribution, e.g.
  \[
  F_E(v) = \frac{\left| \text{number of samples} \leq v \right|}{m} + \sqrt{\log \frac{1}{\delta}}
  \]
Lower Confidence (Revenue) Bounds vs. Underestimating Distribution

- Recall the take-home question regarding optimal pricing
  - Different value distributions require different regularization in Lecture I
  - Can we get all upper bounds using the same algorithm?

- Lower Confidence Bound (LCB), e.g., choose $p$ to maximize

$$p \cdot \left( \frac{\left| \text{number of samples } \geq p \right|}{m} - \sqrt{\frac{\log \frac{1}{\delta}}{m}} \right)$$

- Alternatively, consider an underestimation of the value distribution, e.g.

$$F_E(v) = \frac{\left| \text{number of samples } \leq v \right|}{m} + \sqrt{\frac{\log \frac{1}{\delta}}{m}}$$

- Two approaches coincide for pricing...
Lower Confidence (Revenue) Bounds vs. Underestimating Distribution

- Recall the take-home question regarding optimal pricing
  - Different value distributions require different regularization in Lecture I
  - Can we get all upper bounds using the same algorithm?

- Lower Confidence Bound (LCB), e.g., choose $p$ to maximize

$$p \cdot \left( \frac{|\text{number of samples} \geq p|}{m} - \sqrt{\frac{\log \frac{1}{\delta}}{m}} \right)$$

- Alternatively, consider an underestimation of the value distribution, e.g.

$$F_E(v) = \frac{|\text{number of samples} \leq v|}{m} + \sqrt{\frac{\log \frac{1}{\delta}}{m}}$$

- Two approaches coincide for pricing... but not for auctions
Product empirical distribution $E = E_1 \times E_2 \times \cdots \times E_n$
- $E_i$ is uniform distribution over bidder $i$’s value samples
Product empirical distribution $E = E_1 \times E_2 \times \cdots \times E_n$
- $E_i$ is uniform distribution over bidder $i$’s value samples

Dominated product empirical distribution $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$$F_{\bar{E}_i}(v) = F_{E_i}(v) + \sqrt{\frac{F_{E_i}(v)(1 - F_{E_i}(v))}{m}}$$

(simplified incorrect form for illustration)
Data-Driven (Single-Item) Auction via Learning Value Distribution

- Product empirical distribution $E = E_1 \times E_2 \times \cdots \times E_n$
  - $E_i$ is uniform distribution over bidder $i$’s value samples

- Dominated product empirical distribution $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$$F_{\bar{E}_i}(v) = F_{E_i}(v) + \sqrt{\frac{F_{E_i}(v)(1 - F_{E_i}(v))}{m}}$$

(simplified incorrect form for illustration)

- Return Myerson’s optimal auction w.r.t. $E$ or $\bar{E}$
Recap

Two Different Viewpoints

Learnability of Product Distributions

Strong (Revenue) Monotonicity

Further Extensions and Open Questions
Hellinger Distance

\[ H(P, Q) = \frac{1}{\sqrt{2}} \left\| \sqrt{P} - \sqrt{Q} \right\|_2 = \sqrt{\frac{1}{2} \sum_v \left( \sqrt{P(v)} - \sqrt{Q(v)} \right)^2} \]
Hellinger Distance

\[ H(P, Q) = \frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{Q} \|_2 = \sqrt{\frac{1}{2} \sum_v (\sqrt{P(v)} - \sqrt{Q(v)})^2} \]

\[ \square \quad \text{Direct product} \]

\[ 1 - H(P_1 \times \cdots \times P_n, Q_1 \times \cdots \times Q_n)^2 = \prod_{i=1}^{n} (1 - H(P_i, Q_i)^2) \]
Hellinger Distance

\[ H(P, Q) = \frac{1}{\sqrt{2}} \| \sqrt{P} - \sqrt{Q} \|_2 = \sqrt{\frac{1}{2} \sum_v \left( \sqrt{P(v)} - \sqrt{Q(v)} \right)^2} \]

- Direct product

\[ 1 - H(P_1 \times \cdots \times P_n, Q_1 \times \cdots \times Q_n)^2 = \prod_{i=1}^{n} \left( 1 - H(P_i, Q_i)^2 \right) \]

- This implies sub-additivity

\[ H(P_1 \times \cdots \times P_n, Q_1 \times \cdots \times Q_n)^2 \leq \sum_{i=1}^{n} H(P_i, Q_i)^2 \]
Hellinger, Kullback–Leibler, and Total Variation

- Relation to TV

\[ H(P, Q)^2 \leq TV(P, Q) \leq \sqrt{2} \cdot H(P, Q) \]
Hellinger, Kullback–Leibler, and Total Variation

- Relation to TV
  \[ H(P, Q)^2 \leq TV(P, Q) \leq \sqrt{2} \cdot H(P, Q) \]

- Relation to KL
  \[ H(P, Q)^2 \leq KL(P \parallel Q) \]
Relation to TV

\[ H(P, Q)^2 \leq TV(P, Q) \leq \sqrt{2} \cdot H(P, Q) \]

Relation to KL

\[ H(P, Q)^2 \leq KL(P \| Q) \]

Why TV is called total variation distance?
Hellinger, Kullback–Leibler, and Total Variation

- Relation to TV
  \[ H(P, Q)^2 \leq \text{TV}(P, Q) \leq \sqrt{2} \cdot H(P, Q) \]

- Relation to KL
  \[ H(P, Q)^2 \leq \text{KL}(P \parallel Q) \]

- Why TV is called total variation distance?
  - \( P \) and \( Q \) are distributions over \( \mathcal{T} \)
  - \( h : \mathcal{T} \to [0, 1] \) is a function
Hellinger, Kullback–Leibler, and Total Variation

- Relation to TV
  \[ H(P, Q)^2 \leq TV(P, Q) \leq \sqrt{2} \cdot H(P, Q) \]

- Relation to KL
  \[ H(P, Q)^2 \leq KL(P || Q) \]

- Why TV is called total variation distance?
  - \( P \) and \( Q \) are distributions over \( \mathcal{T} \)
  - \( h : \mathcal{T} \rightarrow [0, 1] \) is a function
  - We have
    \[ |E_{v \sim P} h(v) - E_{v \sim Q} h(v)| \leq TV(P, Q) \]
Learnability of Distribution

Theorem

If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\epsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \epsilon$$
Learnability of Distribution

Theorem

If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

□ Here we prove a weaker result $E H(D, E) \lesssim \sqrt{\frac{k}{m}}$

$$\left( E H(D, E) \right)^2 \leq E H(D, E)^2$$
Learnability of Distribution

Theorem

If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

Here we prove a weaker result

$$\mathbb{E} H(D, E) \lesssim \sqrt{\frac{k}{m}}$$

$$\left( \mathbb{E} H(D, E) \right)^2 \leq \mathbb{E} H(D, E)^2 = \sum_v \mathbb{E} \left( \sqrt{D(v)} - \sqrt{E(v)} \right)^2$$
Learnability of Distribution

Theorem

If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

- Here we prove a weaker result $E H(D, E) \lesssim \sqrt{\frac{k}{m}}$

$$\left( E H(D, E) \right)^2 \leq E H(D, E)^2 = \sum_v E \left( \sqrt{D(v)} - \sqrt{E(v)} \right)^2$$

- It suffices to bound $E \left( \sqrt{D(v)} - \sqrt{E(v)} \right)^2$ for any $v$

$$E \frac{(D(v) - E(v))^2}{(\sqrt{D(v)} + \sqrt{E(v)})^2}$$
Learnability of Distribution

**Theorem**

*If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\epsilon^2}$ i.i.d. samples, then*

$$H(D, E) \leq \epsilon$$

- Here we prove a weaker result $\mathbb{E} H(D, E) \lesssim \sqrt{\frac{k}{m}}$

$$\left(\mathbb{E} H(D, E)\right)^2 \leq \mathbb{E} H(D, E)^2 = \sum_v \mathbb{E} \left(\sqrt{D(v)} - \sqrt{E(v)}\right)^2$$

- It suffices to bound $\mathbb{E} \left(\sqrt{D(v)} - \sqrt{E(v)}\right)^2$ for any $v$

$$\mathbb{E} \frac{(D(v) - E(v))^2}{(\sqrt{D(v)} + \sqrt{E(v)})^2} \leq \mathbb{E} \frac{(D(v) - E(v))^2}{D(v)}$$

\(\chi^2\) distance
Learnability of Distribution

Theorem

If $D$ has support size $k$, $E$ is empirical distribution over $m \approx \frac{k + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

- Here we prove a weaker result $E H(D, E) \lesssim \sqrt{\frac{k}{m}}$

$$\left( E H(D, E) \right)^2 \leq E H(D, E)^2 = \sum_v E \left( \sqrt{D(v)} - \sqrt{E(v)} \right)^2$$

- It suffices to bound $E \left( \sqrt{D(v)} - \sqrt{E(v)} \right)^2$ for any $v$

$$E \frac{(D(v) - E(v))^2}{(\sqrt{D(v)} + \sqrt{E(v)})^2} \leq E \frac{(D(v) - E(v))^2}{D(v)} = \frac{1}{m} \left( 1 - D(v) \right) \leq \frac{1}{m}$$

$\chi^2$ distance
Learnability of Product Distribution

Theorem

If $D = D_1 \times D_2 \times \cdots \times D_n$ and each $D_i$ has support size $k$, $E = E_1 \times E_2 \times \cdots \times E_n$ is the product empirical distribution over $m \approx \frac{kn + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$
Learnability of Product Distribution

Theorem
If \( D = D_1 \times D_2 \times \cdots \times D_n \) and each \( D_i \) has support size \( k \), \( E = E_1 \times E_2 \times \cdots \times E_n \) is the product empirical distribution over \( m \approx \frac{kn+\log \frac{1}{\delta}}{\varepsilon^2} \) i.i.d. samples, then

\[
H(D, E) \leq \varepsilon
\]

- Again, we prove a weaker result \( \mathbb{E} H(D, E) \lesssim \sqrt{\frac{kn}{m}} \)
Learnability of Product Distribution

Theorem

If $D = D_1 \times D_2 \times \cdots \times D_n$ and each $D_i$ has support size $k$, $E = E_1 \times E_2 \times \cdots \times E_n$ is the product empirical distribution over $m \approx \frac{kn + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

- Again, we prove a weaker result $E H(D, E) \lesssim \sqrt{\frac{kn}{m}}$

$$\left( E H(D, E) \right)^2 \leq E H(D, E)^2$$
Learnability of Product Distribution

Theorem

If \( D = D_1 \times D_2 \times \cdots \times D_n \) and each \( D_i \) has support size \( k \), \( E = E_1 \times E_2 \times \cdots \times E_n \) is the product empirical distribution over \( m \approx \frac{kn + \log \frac{1}{\delta}}{\varepsilon^2} \) i.i.d. samples, then

\[
H(D, E) \leq \varepsilon
\]

- Again, we prove a weaker result \( \mathbb{E} H(D, E) \lesssim \sqrt{\frac{kn}{m}} \)

\[
\left( \mathbb{E} H(D, E) \right)^2 \leq \mathbb{E} H(D, E)^2 \leq \sum_{i=1}^{n} \mathbb{E} H(D_i, E_i)^2
\]
Learnability of Product Distribution

Theorem

If $D = D_1 \times D_2 \times \cdots \times D_n$ and each $D_i$ has support size $k$, $E = E_1 \times E_2 \times \cdots \times E_n$ is the product empirical distribution over $m \approx \frac{kn + \log \frac{1}{\delta}}{\varepsilon^2}$ i.i.d. samples, then

$$H(D, E) \leq \varepsilon$$

☐ Again, we prove a weaker result $E H(D, E) \lesssim \sqrt{\frac{kn}{m}}$

$$\left( E H(D, E) \right)^2 \leq E H(D, E)^2 \leq \sum_{i=1}^{n} E H(D_i, E_i)^2 \leq \sum_{i=1}^{n} \frac{k}{m}$$
Learnability of Product Distribution

Theorem

If \( D = D_1 \times D_2 \times \cdots \times D_n \) and each \( D_i \) has support size \( k \), \( E = E_1 \times E_2 \times \cdots \times E_n \) is the product empirical distribution over \( m \approx \frac{kn \cdot \log \frac{1}{\delta}}{\varepsilon^2} \) i.i.d. samples, then

\[
H(D, E) \leq \varepsilon
\]

☐ Again, we prove a weaker result \( EH(D, E) \lesssim \sqrt{\frac{kn}{m}} \)

\[
\left( EH(D, E) \right)^2 \leq EH(D, E)^2 \leq \sum_{i=1}^{n} EH(D_i, E_i)^2 \leq \sum_{i=1}^{n} \frac{k}{m} = \frac{kn}{m}
\]
Implications to Data-Driven Auction Design

- $D = D_1 \times D_2 \times \cdots \times D_n$ is an $n$-dimensional product value distribution
Implications to Data-Driven Auction Design

- $D = D_1 \times D_2 \times \cdots \times D_n$ is an $n$-dimensional product value distribution
  - We may think of each dimension’s support size as $k = \frac{1}{\varepsilon}$
    - because we can round values $v_i$ to $\lfloor v_i \rfloor_\varepsilon$ (closest multiple of $\varepsilon$)
Implications to Data-Driven Auction Design

- $D = D_1 \times D_2 \times \cdots \times D_n$ is an $n$-dimensional product value distribution
  - We may think of each dimension’s support size as $k = \frac{1}{\varepsilon}$ because we can round values $v_i$ to $\lfloor v_i \rfloor_\varepsilon$ (closest multiple of $\varepsilon$)
  - Formally, let $\lfloor D \rfloor_\varepsilon$ be the distribution of rounded value profile

\[
\text{OPT}(\lfloor D \rfloor_\varepsilon) \geq \text{OPT}(D) - \varepsilon
\]
Implications to Data-Driven Auction Design

- $D = D_1 \times D_2 \times \cdots \times D_n$ is an $n$-dimensional product value distribution
  - We may think of each dimension’s support size as $k = \frac{1}{\varepsilon}$ because we can round values $v_i$ to $\lfloor v_i \rfloor_\varepsilon$ (closest multiple of $\varepsilon$)
  - Formally, let $\lfloor D \rfloor_\varepsilon$ be the distribution of rounded value profile
    $$OPT(\lfloor D \rfloor_\varepsilon) \geq OPT(D) - \varepsilon$$

- $E = E_1 \times E_2 \times \cdots \times E_n$ is product empirical distribution from $m$ rounded samples
  - $E_i$ is the uniform distribution over bidder $i$’s rounded sample values
Implications to Data-Driven Auction Design

- $D = D_1 \times D_2 \times \cdots \times D_n$ is an $n$-dimensional product value distribution
  - We may think of each dimension’s support size as $k = \frac{1}{\varepsilon}$ because we can round values $v_i$ to $\lfloor v_i \rfloor_\varepsilon$ (closest multiple of $\varepsilon$)
  - Formally, let $\lfloor D \rfloor_\varepsilon$ be the distribution of rounded value profile
    
    $$OPT(\lfloor D \rfloor_\varepsilon) \geq OPT(D) - \varepsilon$$

- $E = E_1 \times E_2 \times \cdots \times E_n$ is product empirical distribution from $m$ rounded samples
  - $E_i$ is the uniform distribution over bidder $i$’s rounded sample values

**Theorem**

With $m \gtrsim \frac{n}{\varepsilon^3} + \frac{\log \frac{1}{\delta}}{\varepsilon^2}$ samples, Myerson’s optimal auction $M_E$ w.r.t. $E$ is an $\varepsilon$ additive approximation w.p. $1 - \delta$. 
Recap

Two Different Viewpoints

Learnability of Product Distributions

Strong (Revenue) Monotonicity

Further Extensions and Open Questions
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples

![Diagram showing CDF](image-url)
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$.
- Empirical distribution $E$ over $m$ samples.
- Bernstein Inequality + Union Bound

\[
|F_E(v) - F_D(v)| \lesssim \sqrt{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta} + \log \frac{m}{\delta}}
\]
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples
- Bernstein Inequality + Union Bound

$$|F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}$$

- Dominated empirical $\bar{E}$

$$F_{\bar{E}}(v) - F_E(v) \approx \sqrt{\frac{F_E(v)(1 - F_E(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}$$
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples
- Bernstein Inequality + Union Bound
  \[
  |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v))}{m} \log \frac{m}{\delta}} + \frac{\log \frac{m}{\delta}}{m}
  \]
- Dominated empirical $\bar{E}$
  \[
  F_{\bar{E}}(v) - F_E(v) \approx \sqrt{\frac{F_E(v)(1 - F_E(v))}{m} \log \frac{m}{\delta}} + \frac{\log \frac{m}{\delta}}{m}
  \]
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples
- Bernstein Inequality + Union Bound
  \[ |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m} \]

- Dominated empirical $\bar{E}$
  \[ F_{\bar{E}}(v) - F_E(v) \lesssim \sqrt{\frac{F_E(v)(1 - F_E(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m} \]

- Auxiliary distribution $\bar{D}$ \quad ($\bar{D} \preceq \bar{E} \preceq D$)
  \[ F_{\bar{D}}(v) - F_D(v) \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m} \]
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution $\bar{E}_i$ for each bidder $i$
- Return Myerson’s optimal auction $M_{\bar{E}}$ w.r.t. $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$M_{\bar{E}}(D)$ vs. $OPT(D)$
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution $\bar{E}_i$ for each bidder $i$
- Return Myerson’s optimal auction $M_{\bar{E}}$ w.r.t. $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$M_{\bar{E}}(D)$ vs. $OPT(D)$

- What’s the best conceivable lower bound for $M_{\bar{E}}(D)$ given $\bar{E} \preceq D$?
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution $\bar{E}_i$ for each bidder $i$
- Return Myerson’s optimal auction $M_{\bar{E}}$ w.r.t. $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$M_{\bar{E}}(D)$ vs. $OPT(D)$

- What’s the best conceivable lower bound for $M_{\bar{E}}(D)$ given $\bar{E} \preceq D$?

$M_{\bar{E}}(D) \geq OPT(\bar{E})$ (strong monotonicity)
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution $\bar{E}_i$ for each bidder $i$
- Return Myerson’s optimal auction $M_{\bar{E}}$ w.r.t. $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

$$M_{\bar{E}}(D) \text{ vs. } OPT(D)$$

- What’s the best conceivable lower bound for $M_{\bar{E}}(D)$ given $\bar{E} \preceq D$?
  $$M_{\bar{E}}(D) \geq OPT(\bar{E}) \quad \text{(strong monotonicity)}$$

- What’s the best conceivable lower bound for $OPT(\bar{E})$ given $\bar{E} \succeq \bar{D}$?
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution \( \bar{E}_i \) for each bidder \( i \)
- Return Myerson’s optimal auction \( M_{\bar{E}} \) w.r.t. \( \bar{E} = \bar{E}_1 \times \bar{E}_2 \times \ldots \times \bar{E}_n \)

\[
M_{\bar{E}}(D) \; \text{vs.} \; OPT(D)
\]

- What’s the best conceivable lower bound for \( M_{\bar{E}}(D) \) given \( \bar{E} \preceq D \)?
  \[
  M_{\bar{E}}(D) \geq OPT(\bar{E}) \quad \text{(strong monotonicity)}
  \]

- What’s the best conceivable lower bound for \( OPT(\bar{E}) \) given \( \bar{E} \succeq \bar{D} \)?
  \[
  OPT(\bar{E}) \geq OPT(\bar{D}) \quad \text{(weak monotonicity)}
  \]
Dominated Empirical Myerson’s Auction

- Compute dominated empirical distribution $\bar{E}_i$ for each bidder $i$
- Return Myerson’s optimal auction $M_{\bar{E}}$ w.r.t. $\bar{E} = \bar{E}_1 \times \bar{E}_2 \times \cdots \times \bar{E}_n$

\[ M_{\bar{E}}(D) \text{ vs. } OPT(D) \]

- What’s the best conceivable lower bound for $M_{\bar{E}}(D)$ given $\bar{E} \preceq D$?
  \[ M_{\bar{E}}(D) \geq OPT(\bar{E}) \]
  (strong monotonicity)

- What’s the best conceivable lower bound for $OPT(\bar{E})$ given $\bar{E} \succeq \bar{D}$?
  \[ OPT(\bar{E}) \geq OPT(\bar{D}) \]
  (weak monotonicity)

- Compare $OPT(\bar{D})$ and $OPT(D)$
Strong (Revenue) Monotonicity

Theorem

For any value distributions $D \succeq \bar{E}$, and the optimal auction $M_{\bar{E}}$ for $\bar{E}$

$$M_{\bar{E}}(D) \geq OPT(\bar{E})$$
Strong (Revenue) Monotonicity

Theorem
For any value distributions $D \succeq \bar{E}$, and the optimal auction $M_{\bar{E}}$ for $\bar{E}$

$$M_{\bar{E}}(D) \geq \text{OPT}(\bar{E})$$

Here we only prove weak monotonicity, i.e., $\text{OPT}(D) \geq \text{OPT}(\bar{E})$, via coupling.
Strong (Revenue) Monotonicity

Theorem

For any value distributions \( D \succeq \bar{E} \), and the optimal auction \( M_{\bar{E}} \) for \( \bar{E} \)

\[
M_{\bar{E}}(D) \geq \text{OPT}(\bar{E})
\]

Here we only prove weak monotonicity, i.e., \( \text{OPT}(D) \geq \text{OPT}(\bar{E}) \), via coupling

quantiles \( q_1, q_2, \ldots, q_n \)
Strong (Revenue) Monotonicity

Theorem

For any value distributions $D \succeq \bar{E}$, and the optimal auction $M_{\bar{E}}$ for $\bar{E}$

$$M_{\bar{E}}(D) \geq OPT(\bar{E})$$

Here we only prove weak monotonicity, i.e., $OPT(D) \geq OPT(\bar{E})$, via coupling

quantiles $q_1, q_2, \ldots, q_n$

1. Values $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n \sim \bar{E}$
2. Allocate to bidder $i$ with highest non-negative $\bar{\phi}_{\bar{E}_i}(\bar{v}_i)$
3. Winner pays threshold bid
Strong (Revenue) Monotonicity

Theorem

For any value distributions \( D \succeq \bar{E} \), and the optimal auction \( M_{\bar{E}} \) for \( \bar{E} \)

\[
M_{\bar{E}}(D) \geq OPT(\bar{E})
\]

Here we only prove weak monotonicity, i.e., \( OPT(D) \geq OPT(\bar{E}) \), via coupling

quantiles \( q_1, q_2, \ldots, q_n \)

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

1. Values \( \bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n \sim \bar{E} \)
Strong (Revenue) Monotonicity

Theorem
For any value distributions $D \succeq \bar{E}$, and the optimal auction $M_{\bar{E}}$ for $\bar{E}$

$$M_{\bar{E}}(D) \geq \text{OPT}(\bar{E})$$

Here we only prove weak monotonicity, i.e., $\text{OPT}(D) \geq \text{OPT}(\bar{E})$, via coupling quantiles $q_1, q_2, \ldots, q_n$

1. Values $v_1, v_2, \ldots, v_n \sim D$
2. Allocate to bidder $i$ with highest non-negative $\bar{\phi}_{E_i}(v_i)$

1. Values $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n \sim \bar{E}$
2. Allocate to bidder $i$ with highest non-negative $\bar{\phi}_{E_i}(\bar{v}_i)$
3. Winner pays threshold bid
Strong (Revenue) Monotonicity

Theorem

For any value distributions $D \succeq \bar{E}$, and the optimal auction $M_{\bar{E}}$ for $\bar{E}$

$$M_{\bar{E}}(D) \geq \text{OPT} (\bar{E})$$

Here we only prove weak monotonicity, i.e., $\text{OPT} (D) \geq \text{OPT} (\bar{E})$, via coupling quantiles $q_1, q_2, \ldots, q_n$

1. Values $v_1, v_2, \ldots, v_n \sim D$
2. Allocate to bidder $i$ with highest non-negative $\bar{\varphi}_{\bar{E}_i}(\bar{v_i})$
3. Winner pays threshold bid, which is at least as large

1. Values $\bar{v}_1, \bar{v}_2, \ldots, \bar{v}_n \sim \bar{E}$
2. Allocate to bidder $i$ with highest non-negative $\bar{\varphi}_{\bar{E}_i}(\bar{v_i})$
3. Winner pays threshold bid
Comparing $OPT(D)$ and $OPT(\bar{D})$

**Lemma**

If we have $m \gtrsim \frac{n \cdot (\log \frac{m}{\varepsilon \delta})^2}{\varepsilon^2}$ samples, then the auxiliary distribution $\bar{D}$

$$H(D, \bar{D}) \leq \frac{\varepsilon}{\sqrt{2}}$$

**Reminder**

$$F_{\bar{D}}(v) - F_D(v) \approx \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}$$
Comparing $OPT(D)$ and $OPT(\bar{D})$

$\bar{D}(v) - D(v) \approx \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}$

Reminder

Lemma

If we have $m \gtrsim n \cdot (\log \frac{m}{\varepsilon \delta})^2$ samples, then the auxiliary distribution $\bar{D}$

$$H(D, \bar{D}) \leq \frac{\varepsilon}{\sqrt{2}}$$

$$\Rightarrow$$

$$TV(D, \bar{D}) \leq \varepsilon$$
Comparing $OPT(D)$ and $OPT(\bar{D})$

Lemma

If we have $m \gtrsim \frac{n \cdot (\log \frac{m}{\epsilon \delta})^2}{\epsilon^2}$ samples, then the auxiliary distribution $\bar{D}$

$$H(D, \bar{D}) \leq \frac{\epsilon}{\sqrt{2}}$$

$$\Rightarrow \quad TV(D, \bar{D}) \leq \epsilon$$

$$\Rightarrow \quad \text{for any auction } A, \quad A(\bar{D}) \geq A(D) - \epsilon$$
Comparing $OPT(D)$ and $OPT(\bar{D})$

Lemma

If we have $m \gtrsim n \cdot (\log \frac{m}{\epsilon\delta})^2$ samples, then the auxiliary distribution $\bar{D}$

$$H(D, \bar{D}) \leq \frac{\epsilon}{\sqrt{2}}$$

$$\Rightarrow$$

$$TV(D, \bar{D}) \leq \epsilon$$

$$\Rightarrow$$

for any auction $A$, $A(\bar{D}) \geq A(D) - \epsilon$

$$\Rightarrow$$

$$OPT(\bar{D}) \geq OPT(D) - \epsilon$$
**Summary**

<table>
<thead>
<tr>
<th>Distributions</th>
<th>Sample Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>[0, 1]-Bounded</td>
<td>$\frac{n}{\varepsilon^2}$</td>
</tr>
<tr>
<td>Regular distributions</td>
<td>$\frac{n}{\varepsilon^3}$</td>
</tr>
<tr>
<td>MHR distributions</td>
<td>$\frac{n}{\varepsilon^2}$</td>
</tr>
<tr>
<td>[1, $H$]-bounded distributions</td>
<td>$\frac{Hn}{\varepsilon^2}$</td>
</tr>
</tbody>
</table>

- **Upper Bound:**
  Learnability of product distribution + strong (revenue) monotonicity

- **Lower Bound:**
  Assouad’s method
Recap

Two Different Viewpoints

Learnability of Product Distributions

Strong (Revenue) Monotonicity

Further Extensions and Open Questions
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization

- Single-parameter auctions (e.g., multiple homogeneous items)
- Multi-parameter auctions (e.g., multiple heterogeneous items)

- Optimal sample complexity is still open

- Sequential decision-making in stochastic models
- Prophet inequality
- Optimal sample complexity is still open
- Pandora's box
- Online stochastic matching
- Optimal sample complexity is still open
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
  - Multi-parameter auctions (e.g., multiple heterogeneous items)
    Optimal sample complexity is still open
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
  - Multi-parameter auctions (e.g., multiple heterogeneous items)
    Optimal sample complexity is still open

- Sequential decision-making in stochastic models
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
  - Multi-parameter auctions (e.g., multiple heterogeneous items)
    Optimal sample complexity is still open

- Sequential decision-making in stochastic models
  - Prophet inequality
    Optimal sample complexity is still open
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
  - Multi-parameter auctions (e.g., multiple heterogeneous items)
    Optimal sample complexity is still open

- Sequential decision-making in stochastic models
  - Prophet inequality
    Optimal sample complexity is still open
  - Pandora’s box
Sample Complexity of Optimization Problems in Stochastic Models

- Revenue maximization
  - Single-parameter auctions (e.g., multiple homogeneous items)
  - Multi-parameter auctions (e.g., multiple heterogeneous items)
    - Optimal sample complexity is still open

- Sequential decision-making in stochastic models
  - Prophet inequality
    - Optimal sample complexity is still open
  - Pandora’s box
  - Online stochastic matching
    - Optimal sample complexity is still open
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples
- Bernstein Inequality + Union Bound

$$|F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v))\log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}$$
Underestimating Value Distribution

- Value distribution $D$, e.g., uniform on $[0, 1]$
- Empirical distribution $E$ over $m$ samples
- Bernstein Inequality + Union Bound
  \[ |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m} \]
- Dvoretzky–Kiefer–Wolfowitz (DKW) inequality
  \[ |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{\log \frac{1}{\delta}}{m}} \]
Underestimating Value Distribution

- **Value distribution $D$, e.g., uniform on $[0, 1]$**
- **Empirical distribution $E$ over $m$ samples**
- **Bernstein Inequality + Union Bound**
  \[
  |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{F_D(v)(1 - F_D(v)) \log \frac{m}{\delta}}{m}} + \frac{\log \frac{m}{\delta}}{m}
  \]
- **Dvoretzky–Kiefer–Wolfowitz (DKW) inequality**
  \[
  |F_E(v) - F_D(v)| \lesssim \sqrt{\frac{\log \frac{1}{\delta}}{m}}
  \]
- **Open question:** Is there a Bernstein-style DKW inequality?
Bidders’ Strategic Behaviors in Data-Driven Auction Design

- Bidders may underbid today in order to get a lower price tomorrow.
Bidders’ Strategic Behaviors in Data-Driven Auction Design

- Bidders may underbid today in order to get a lower price tomorrow
- Can we learn optimal auctions despite of bidders’ strategic behaviors?
Bidders’ Strategic Behaviors in Data-Driven Auction Design

- Bidders may underbid today in order to get a lower price tomorrow
- Can we learn optimal auctions despite of bidders’ strategic behaviors?
  - Impossible if bidders are patient
Bidders’ Strategic Behaviors in Data-Driven Auction Design

- Bidders may underbid today in order to get a lower price tomorrow
- Can we learn optimal auctions despite of bidders’ strategic behaviors?
  - Impossible if bidders are patient
  - Possible for relatively simple auctions, and impatient bidders
    (with slower convergence rate than learning form non-strategic bidders)
Bidders’ Strategic Behaviors in Data-Driven Auction Design

- Bidders may underbid today in order to get a lower price tomorrow.
- Can we learn optimal auctions despite of bidders’ strategic behaviors?
  - Impossible if bidders are patient.
  - Possible for relatively simple auctions, and impatient bidders (with slower convergence rate than learning form non-strategic bidders).
  - **Open question:** Is the slower convergence rate avoidable?
References

