Robust Inference for Games via Theoretical Guarantees

Denis Nekipelov
(with Darrell Hoy and Vasilis Syrgkanis)

Simons Institute Conference:
Algorithmic Game Theory and Practice

November 18

Structural inference

- **Main Goal of Structural Inference:** use data to understand complex systems . . . so that good decision can be made (i.e.,optimizing performance).
- **Assumption:** agents in the system respond to its design and each other (and we can model how)
- **Fundamental Challenge:** Need to predict behavior in new system from behavior in old system
- **Observation:** The "state of the world" for the new design may not be observable

Conclusion: Need to recover the "primitives" of the model (e.g. preferences of agents) to make predictions

- **This Talk:** Counterfactual analysis for auctions

Practical questions

- Questions of "counterfactual predictions"
  - Compare mechanism A applied to a given population with mechanism B (e.g. optimal mechanism)
  - Propose the direction for improvement of the mechanism

- Usual requirements to answers (in digital platforms)
  - Computationally efficient
  - Scalable
  - Robust to modeling assumptions (prefer tuning parameter-free)

Equilibrium and Inference

- **Assumption:** bidders are happy with their bids.
- **Equilibrium:** bidder's bid must be best response to competing bid distribution.

Observation:

competing bids distribution is observed in data.

- **Approach:**
  1. given bid distribution, solve for bid strategy
  2. invert bid strategy to get bidder's value for item from bid.

Inference via identification

- This approach is dominant in the "structural inference" literature
  1. Ensure that the model is identified
  2. We use the data to infer the primitives
  3. Then we use the inferred primitives for predictions (a.k.a. counterfactuals)
- There are serious caveats
  - Many structural models are exactly identified. Even worse, many are identified at infinity.
  - This leads to problems with inference
  - Even more serious problems with counterfactuals

Inference for First-price Auction

Utility of the bidders:

for first price auction with allocation rule x and q-quantile of values

$$U(q; x) = x(q)(v(q) - b(q))$$

Notes:

- allocation rule $x(\cdot)$ is determined by the auction mechanism.
- action space determines the bid function $q \mapsto b$ (quantile function of bid distribution).
- monotonicity ensures that $b \leftrightarrow v$

Inference for First-price Auction

Inference Equation:

for first price auction q-quantile of values

$$v(q) = b(q) + \frac{x(q)b'(q)}{x'(q)}$$

Notes:

- bid function $b(\cdot)$, $b'(\cdot)$ must be inferred.
- value function $v(\cdot)$ can be inferred from $\widehat{v}(q) = \widehat{b}(q) + \frac{x(q)\widehat{b}'(q)}{x'(q)}$.
- In i.i.d. setting observe N samples from $b(\cdot)$
- Once we have values, we can predict behavior in the new platform

- $b(\cdot)$ is inferred directly from order statistics
- $b'(\cdot)$ is significantly more problematic

Inference for First-price Auction

Inference Equation:

for first price auction q-quantile of values

$$v(q) = b(q) + \frac{x(q)b'(q)}{x'(q)}$$

Notes:

- Note that $b'(q)$ is inverse of bid density
- If $f_b(b) = e^{-b}$, then

$$\mathrm{Var}(b'(q)) \propto \int \frac{1}{f_b(b)}\, db \to \infty$$

- Standard "concentration results" (Chebychev's inequality, Central Limit theorem) do not apply
  - Inference requires very large samples
  - Results can be non-robust to "local" deviations

Inference for First-price Auction

Inference Equation:

for first price auction q-quantile of values

$$v(q) = b(q) + \frac{x(q)b'(q)}{x'(q)}$$

Notes:

- Non-robustness means that (without additional constraints) there exist close distributions of bids (in the L_∞ norm) that lead to arbitrarily different couterfactual revenue or welfare predictions
- It will not be “fixed” if bids have bounded support

Solution:

- Consider the *entire set* of predictions for all distributions that are considered close
- Can we mainstream process by avoiding computation of all possible values?

Set inference and partial identification

- Manski (1988, 1990) proposes to consider *all* models that could have generated the data
- **Example:** Survey non-response
  - Object of interest is the expectation of outcome $Y \in [Y_L,\ Y_U]$
  - Can take a "good" subsample of population, but there is significant non-response
  - $D = 1$ if responded, $D = 0$ if not (and both can be correlated with Y)
  - Observe probability $P(D = 1)$ and $E[Y|D = 1]$ (outcome for survey responders)

$$\begin{aligned} E[Y] &= E[Y|D=1]P(D=1) + E[Y|D=0](1-P(D=1)) \\ &\in [E[Y|D=1]P(D=1) + Y_L(1-P(D=1)), \\ &\qquad\qquad E[Y|D=1]P(D=1) + Y_U(1-P(D=1))] \end{aligned}$$

  - This bound cannot be improved without additional information on the distribution of Y
  - This is *identified set* for $E[Y]$

Set inference and partial identification

- More generally, we can consider economic model characterized by a vector of parameters $\theta \in \Theta$
- Observable (in the data) variable $Y \sim F_Y$ and unobservable variable $\epsilon \sim F_\epsilon$
- Functions m_i and m_e link the distributions and parameters:

$$E_{Y,\epsilon}[m_i(Y,\epsilon;\theta)] \leq 0, \quad E_{Y,\epsilon}[m_e(Y,\epsilon;\theta)] = 0. \quad (*)$$

- The set of parameters Θ_I compatible with (*) for a given distribution F_Y is the *identified set* for θ
  - Note that the notion of identified set applies to "population" (i.e. works with entire distribution F_Y rather than a sample from it)
  - With data F_Y is approximated by empirical distribution

Inference for market outcomes

- Identified set Θ_I produces the sets of “primitives” of the game (values or sets of values of players in auction)
- Θ_I itself may not be of ultimate interest
- E.g. the goal of structural inference can be to produce counterfactual for “aggregate objects,” e.g. actual vs optimal welfare
- That requires projection of Θ_I on some small subspace

Inference for market outcomes

- In game-theoretic settings the object of interest is the outcome of the counterfactual mechanism (e.g. optimal auction)
- If we have the identified set for the primitives Θ_I, we can compute the counterfactual outcome as a new equilibrium
- When we have sets characterizing preferences, inference becomes very hard
  - Need to infer sets of possible outcomes for each possible value of preferences
  - Even harder if agents' behavior deviates from Nash
- It is preferable to have approach that generates the set of outcomes directly (bypassing the computation of Θ_I).

Constructing bounds for welfare

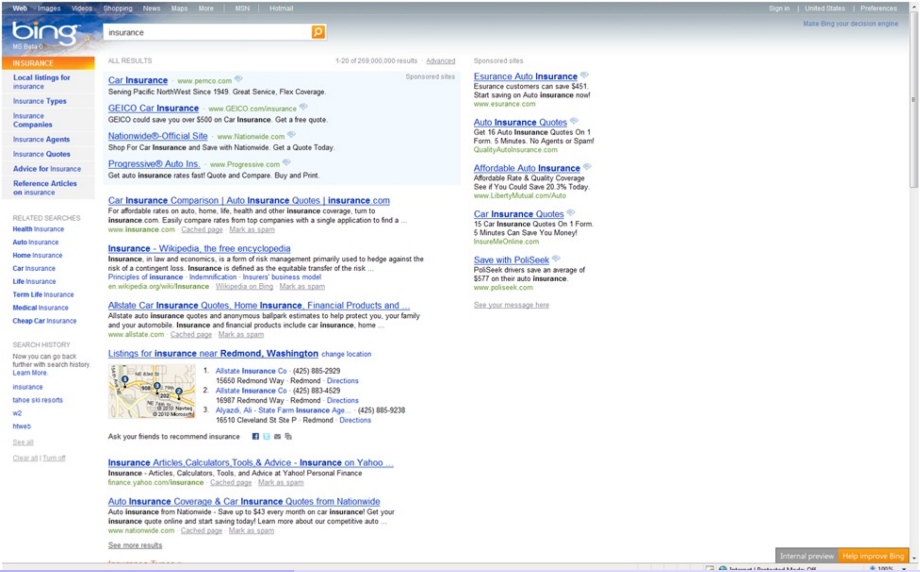
- Koutsoupias, Papadimitrou (1999) introduce the notion of *price of anarchy* (PoA)
- PoA is "worst case" ratio of welfare of given mechanism to optimal welfare (for all considered value distributions and actions of agents)
- Derivation of PoA is based on unilateral deviations of bidders from stable outcomes
- It was found theoretically useful: many common mechanisms have small PoA for large classes of value distributions
- In simulations, however, actual welfare ratio can be substantially smaller than theoretical PoA

Constructing bounds for welfare

- In the "worst-case scenario" PoA is interpreted as a property of the mechanism
  - PoA bounds welfare over all possible distributions of values
- Not all values can occur with equal probabilities
- Knowledge of distribution of actions (bids) imposes implicit constraint on possible distributions of values
- PoA *subject to* observed distribution of bids (call it Empirical PoA or EPoA) produces "realistic" welfare bounds
- *EPoA is the combined property of the distribution of values and the mechanism*
- Closely related to notion of identified set

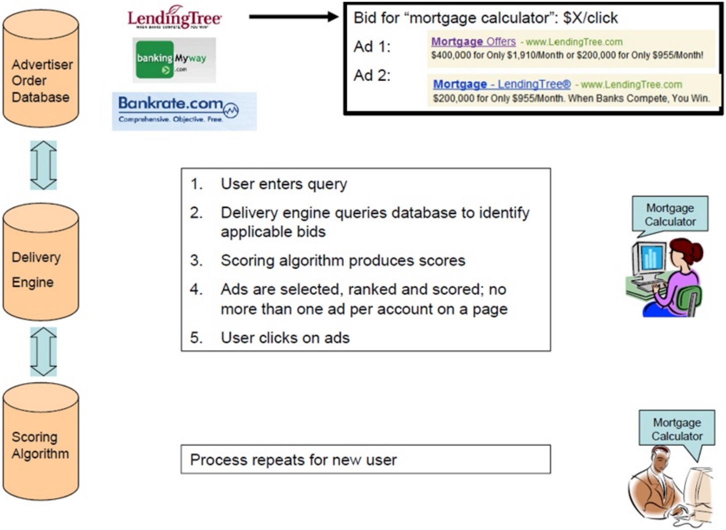
Application

Search ads monetize consumer searches on the Internet



- The ads are allocated and priced for *each user query*
- Pricing and allocation mechanisms are combined and fully automated by an "auction":
  - Real-time
  - Pay per click
  - Score-weighted
  - Generalized second price (GSP)
  - With possible reserve prices and thresholds

Allocation and pricing heterogenous objects


Advertiser Order Database
Delivery Engine
Scoring Algorithm
LendingTree
banking Myway .com
Bankrate.com
Comprehensive. Objective. Free.
Bid for "mortgage calculator": $X/click
Ad 1:
Mortgage Offers - www.LendingTree.com
$400,000 for Only $1,910/Month or $200,000 for Only $955/Month!
Ad 2:
Mortgage - LendingTree® - www.LendingTree.com
$200,000 for Only $955/Month. When Banks Compete, You Win.
1. User enters query
2. Delivery engine queries database to identify applicable bids
3. Scoring algorithm produces scores
4. Ads are selected, ranked and scored; no more than one ad per account on a page
5. User clicks on ads
Process repeats for new user
Mortgage Calculator
Mortgage Calculator

- Search queries arrive multiple times per minute for top keywords
- Advertisers do not receive feedback from each query and respond to aggregate clicks and revenues in the stream of queries
- We characterize each advertiser i by a single parameter v_i, her *value per click* (VPC)
- Advertiser's expected profit:

$$\text{Utility}_i(b_i, v_i; \mathbf{b}_{-i}) = v_i \times \text{Expected Clicks}_i(b_i) - \text{Expected Cost}_i(b_i)$$

- In a complete full information NE model we can recover values for each bidder
  - Given competing bids, bidder i buys clicks until the cost of extra click exceeds value
  - Value per click is equal to the marginal cost per click at actual bid
  - Marginal cost per click can be recovered from the data
  - Two key requirements: bidders best response and best response is unique

- Threats to the model
  - Non-monotonicities in marginal cost
  - "Flat spots" in click function
  - Deviation from best responding (i.e. ϵ-best response)
  - Drifting distribution of uncertainty (e.g. changing traffic to the search platform, seasonal effects, etc.)

- In the context of partial identification, we need to consider each issue separately
- That leads to the set of values that we produce for each bidder
- Since uncertainty is correlated across bidders, need to construct joint sets of values
- Typical search phrases contain thousands of eligible ads
- Construction of joint multidimensional sets of values seems excessive if the final goal is welfare or revenue
- PoA approach can address this issue

Constructing bounds for welfare

- Allow variable uncertainty parametrized by θ
- Each θ corresponds to a different distribution of uncertainty (scores, reserve prices, etc.)

Definition

The Bayesian Empirical Price of Anarchy (EPoA) of the sponsored search auction mechanism A is defined as

$$\text{EPoA}(A) = \sup_{v \in \mathcal{V}, \sigma \in \Sigma, \theta \in \Theta} \frac{E_\theta\left[W(OPT, v, \sigma)\right]}{E_\theta\left[W(A, v, \sigma)\right]},$$

such that $P\left(\sigma_i(\theta; v) \leq b\right) = F_b(b)$, where $\mathcal{V}$ is the set of all values, Σ is the set of all considered strategies, Θ is the set of all distributions of uncertainty and $F_b(\cdot)$ is the distribution of bids.

EPoA

- The constrained optimization problem of EPoA may not always be easily solvable
- It turns out, EPoA can be computed “independently” from auction revenue
- Use idea in Hartline, Hoy and Taggart (2014).

Definition

For the price per click ppc as a function of bid

$$\tau_i(z) = \min_{b \mid \text{Clicks}_i(b) \geq z} \{\text{ppc}(b)\} \tag{1}$$

the *threshold* for agent i and average probability of click Q is

$$T_i(Q) = \int_0^Q \tau_i(z)\, dz \tag{2}$$

Revenue covering approach

Definition (Revenue Covering)

Strategy profile σ of auction A is μ-revenue covered if for any feasible allocation $\mathbf{Q}$,

$$\mu \textsc{Rev}(A(\sigma)) \geq \sum_{\sum_i Q_i = \mathbf{Q}} T_i(Q_i). \quad (3)$$

Definition

Auction A is μ-revenue covered if for *any* strategy profile σ, σ and A are μ-revenue covered.

Revenue covering approach

Lemma (Value Covering)

For any bidder i with value v_i and allocation amount Q_i,

$$Utility_i(v_i, \sigma) + \frac{1}{\mu} T_i(Q_i) \geq \frac{1 - e^{-\mu}}{\mu} Q_i v_i. \tag{4}$$

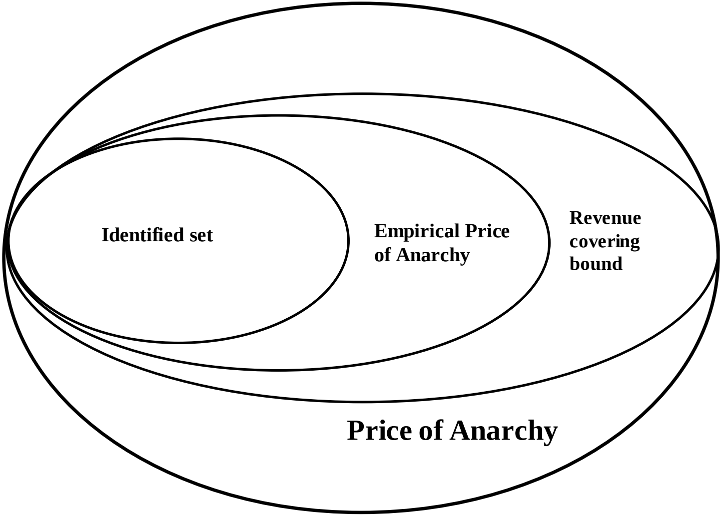
Theorem

The welfare in any μ-revenue covered strategy profile σ of auction A is at least a $\frac{\mu}{1-e^{-\mu}}$-approximation to the optimal welfare.

EPoA

- Revenue covering approach provides the upper bound for the EPoA
- Note that our analysis does not require the auction to be revenue covered in theory
- If the distribution of bids is compatible with some μ for revenue covering, we can apply that μ to bound EPoA
- There is no explicit guarantee that revenue covering approach is tight for EPoA
- There is also no explicit guarantee that EPoA is tight to produce identified set for optimal welfare

Bounds for welfare



EPoA implementation

- Now we have a clear empirical strategy to compute the bounds that we need
  1. For a given mechanism compute threshold functions. There will be one function per bidder. If the mechanism is fully known these function can be computed precisely
  2. Maximize the sum of thresholds over allocations
  3. Compute auction revenue from the data and revenue covering parameter μ
  4. Produce EPoA
- Very attractive from statistical viewpoint: only need empirical revenue

Optimization of thresholds

- To compute revenue covering, need to compute

$$\max_{\sum_i Q_i = \mathbf{Q}} T_i(Q_i)$$

for arbitrary convex $T_i(\cdot)$

- It is NP-hard by a reduction from the maximum hypergraph matching problem (when scores have discrete support)
- Let maximum allocation (maximum possible clicks for bidder i) be

$$\bar{Q}_i = \max_{\mathbf{Q}} Q_i$$

- By convexity of $T_i(\cdot)$:

$$\mathbf{T}^1 = \max_{\mathbf{Q}} \sum_i Q_i \frac{T_i(\bar{Q}_i)}{\bar{Q}_i} \geq \max_{\mathbf{Q}} \sum_i T_i(Q_i)$$

Optimization of thresholds

- Problem is equivalent to welfare maximization where player i has a value-per-click of $\bar{v}_i = \frac{T_i(\bar{Q}_i)}{\bar{Q}_i}$
- Optimal allocation is greedy allocation which ranks bidders by $\text{score}_i \cdot \bar{v}_i$.
- Computing $\mathbf{T}^1$ consists of running a greedy allocation algorithm for each support point of quality scores
- With data,compute optimal greedy allocation for each observed instance of quality scores

Refinement of the bound

Definition (Empirical Value Covering)

Auction A and strategy profile σ are empirically λ-value covered if A is μ-revenue covered, and for any bidder i with value v_i and allocation amount Q_i,

$$\text{Utility}_i(v_i) + \frac{1}{\mu} T_i(Q_i) \geq \frac{\lambda}{\mu} Q_i v_i. \tag{5}$$

Lemma

If auction A and strategy profile σ are empirically μ-revenue covered and λ-value covered, then the empirical price of anarchy of A and σ is at most $\frac{\mu}{\lambda}$.

Refinement of the bound

Lemma

For a μ-revenue covered strategy profile σ and auction A with maximum feasible probabilities of allocation $\overline{x_i}$, let $\underline{\lambda_i^\mu} = \min_{v_i, Q_i'} \frac{\mu u_i(v_i) + T_i(x_i')}{Q_i' v_i}$ and $\underline{\lambda^\mu} = \min_i \underline{\lambda_i^\mu}$. Then A and σ are empirically $\underline{\lambda^\mu}$-value covered.

- If auction is μ-revenue covered w.r.t. $\mathbf{T}^1$, only consider the allocation amount $\bar{Q}^i$
- Optimization reduces to

$$\underline{\lambda_i^\mu} = \min_{v_i} \frac{\mu u_i + T_i(\bar{Q}_i)}{v_i}$$

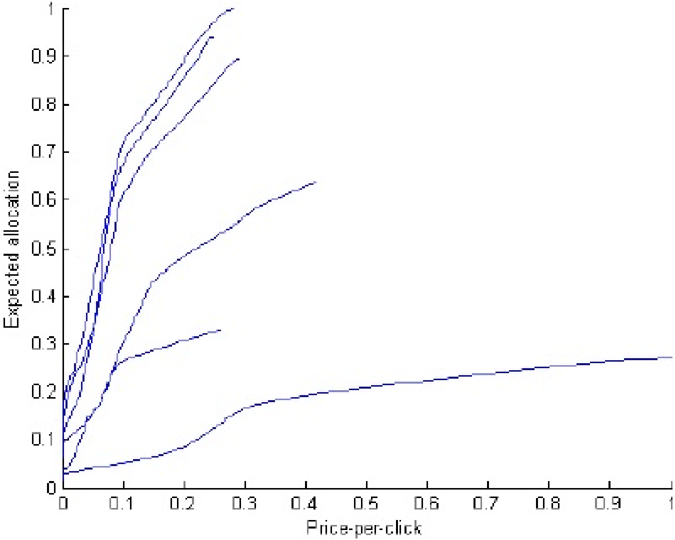
(Statistical) rate optimality

- Recall that problems of inversion of best responses lead to non-existence of second moments
- Statistically this means that if convergence in distribution occurs, it is
  - (a) Slow (standard deviation down to $O((\log\log N)^{-1})$)
  - (b) The limit is non-standard and depends on unobserved tail behavior
- Our implementation of EPoA is robust since it only requires computing revenue and thresholds
- Both can be estimated at fast rate (standard deviation is guaranteed to be $O(N^{-1/2})$)
- The limit distribution is Gaussian

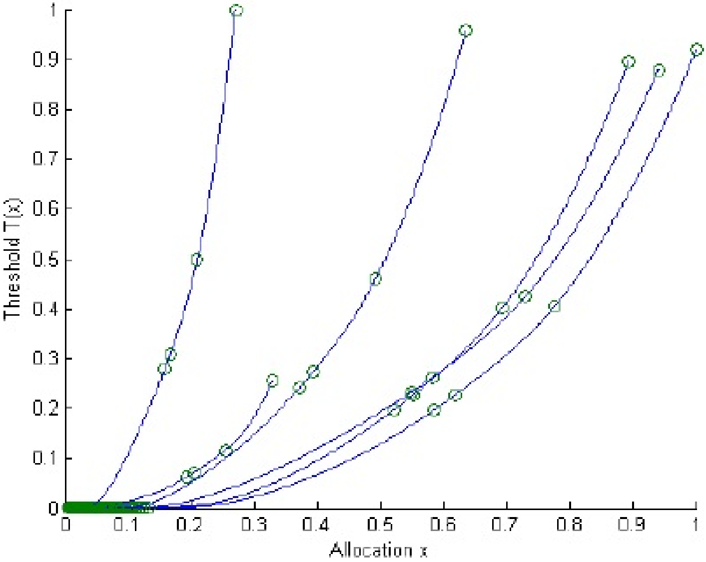
Empirical illustration

- Use historical data from 2014 from advertising platform on Bing.com
- Full access to bidding histories, scoring mechanism, reserve prices
- Select 11 “isolated,” high revenue search phrases
- Use ultra-high frequency bidders (average 2 minutes before bid changes)
- Observe actual tools used by the bidders
- Isolate the period of the week and simulate the components of bidder’s objectives

Empirical illustration



Empirical illustration



Empirical illustration

	$\frac{1}{EPoA^1}$	$\frac{\overline{\mathbf{T}}^1}{\text{Rev}}$	λ^1	$\frac{LB-\mathbf{T}}{\text{Rev}}$	$\frac{1}{LB-EPoA}$	$\frac{\mathbf{T}_{avg}}{\text{Rev}}$	$\frac{1}{FA-EPoA}$
phrase1	.567	1.4164	.803	.562	.788	.511	.783
phrase2	.606	1.2848	.779	.553	.792	.509	.784
phrase3	.279	4.182	1.167	3.364	.325	2.966	.320
phrase4	.284	3.860	1.098	2.298	.401	1.556	.507
phrase5	.673	1.099	.740	.437	.828	.386	.829
phrase6	.628	1.183	.743	.495	.832	.488	.791
phrase7	.672	1.031	.693	.503	.824	.459	.802
phrase8	.645	1.036	.669	.520	.795	.419	.817
phrase9	.622	1.169	.726	.581	.759	.441	.809
phrase10	.597	1.138	.680	.545	.771	.377	.833
phrase11	.573	1.431	.820	.631	.780	.502	.786

Summary

- Focus on construction of counterfactual predictions in structural models
- Traditional approach to such predictions requires the recovery of model primitives from the data
- Equilibrium framework requires functional inversions that can lead to non-robust results
- Partial identification approach produces entire sets of model parameters that are compatible with data and thus more robust
- The price of anarchy approach allows us to consider the inference on the counterfactual outcomes directly
- We develop empirical price of anarchy which is price of anarchy bound derived for all models compatible with observable action distributions
- EPoA is a combined property of the preferences and the mechanism
- The bounds produced by EPoA are closely related with the identified sets considered in Econometrics