High Dimensional Calibration For Rational Decision Making

Aaron Roth
Penn

Joint work with

George Noarov, Ramya Ramalingam, Stephan Xie

Prediction for Action

- Consider a decision maker with action space A and utility function:

$$u: A \times S \to [0,1]$$

for some *state space* $S \subset \mathbb{R}^d$. $u(a,\cdot)$ is linear, Lipschitz in s for all a.

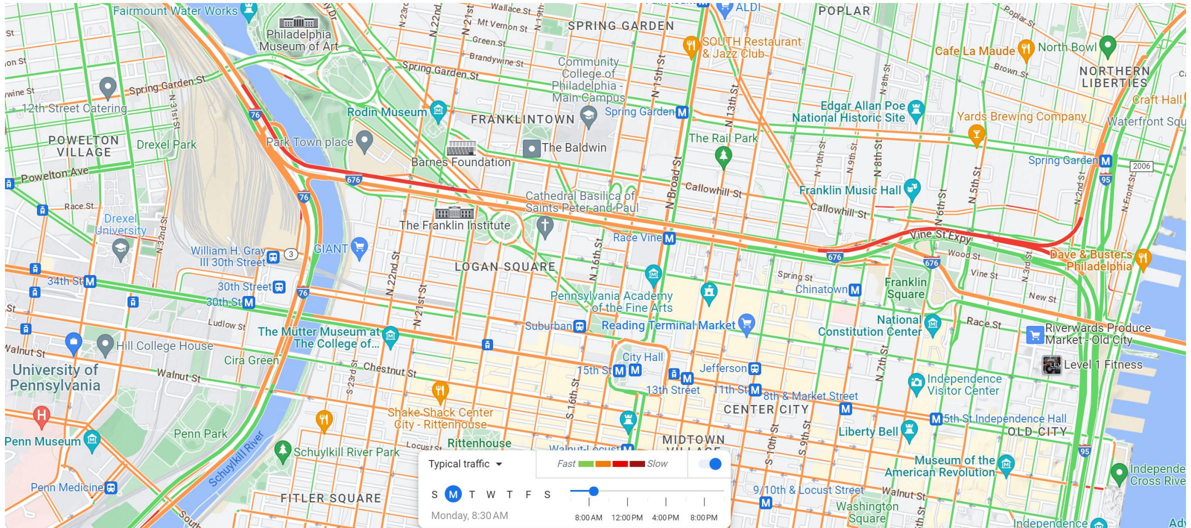
- Models, e.g.
  - A decision maker with *arbitrary* utilities $u(a, o)$ for d distinct outcomes;
    - s is a distribution over outcomes.
  - Routing games (and other congestion games);
    - s is a vector of road/edge congestions
  - Extensive form games
    - s is a vector of probabilities that each terminal node is made reachable by opponents
  - …

Prediction for Action

- If s were known, action would be easy. Could just *best respond.*

$$BR(u, s) = \arg\max_{a \in A} u(a, s)$$

- Difficulty: Must act *before* s is known.
- Suppose we have a *prediction* $\hat{s}$.
- Is it a good idea to play $BR(u, \hat{s})$?



Calibration

- Forecasts $\hat{s}$ are *calibrated* if they are unbiased conditional on their own predictions:

$$\mathbb{E}[s \mid \hat{s}] = \hat{s}$$

- It is a good idea to follow calibrated forecasts.

Theorem: If forecasts $\hat{s}$ are calibrated, then for every u, the best response policy $f^*(\hat{s}) = BR(u, \hat{s})$ is a dominant strategy amongst all policies $f: S \to A$ mapping forecasts to actions.

Calibration

- **Theorem**: If forecasts $\hat{s}$ are calibrated, then for every u, the best response policy $f^*(\hat{s}) = BR(u, \hat{s})$ is a dominant strategy amongst all policies $f: S \to A$ mapping forecasts to actions.

Proof:

$$\begin{aligned}
\mathbb{E}_{s,\hat{s}}[u(f^*(\hat{s}), s)] &= \mathbb{E}_{\hat{s}}\big[\mathbb{E}_s[u(BR(u,\hat{s}), s) | \hat{s}]\big] \\
&= \mathbb{E}_{\hat{s}}[u(BR(u,\hat{s}), \mathbb{E}[s|\hat{s}])] \\
&= \mathbb{E}_{\hat{s}}[u(BR(u,\hat{s}), \hat{s})] \\
&\geq \mathbb{E}_{\hat{s}}[u(f(\hat{s}), \hat{s})]
\end{aligned}$$

Calibration

- Good news:
  - Calibrated predictions incentivize agents to treat them as correct.
  - It is possible to produce calibrated predictions *even for an adversarially chosen sequence of states* [Foster Vohra]
- Unfortunately…
  - One way to achieve calibration is to be uninformative:
    - The constant prediction $\hat{s} = \mathbb{E}[s]$ is calibrated --- but not useful.
  - The number of possible predictions $\hat{s}$ grows exponentially in d
    - The computational and statistical complexity of achieving calibration against adversarial sequences also grows exponentially with d…

This Talk

- In high dimensional settings, what can we ask for short of calibration that still makes it a good idea for agents to treat forecasts as correct?
- Can we obtain it *efficiently* (computationally and statistically) in adversarial settings, and
- Does this lead us to new learning algorithms for large action spaces?
- Does this lead us to new/robust approaches to online mechanism design problems?

The Online Prediction Setting

- A context space X
  - Features relevant to the prediction task
- A convex prediction/outcome space $C \subset \mathbb{R}^d$
  - E.g. the probability simplex, the set of all feasible road congestions, …
- In rounds $t = 1, \ldots, T$:
  - The learner observes some context $x_t \in X$.
  - The learner produces a prediction $\hat{s}_t \in C$
  - The learner observes outcome $s_t \in C$

Making Unbiased Predictions

- An *event* $E(x_t, \hat{s}_t)$ is a function $E: X \times C \to \{0,1\}$. It selects a subsequence of rounds as a function of the context and prediction.
  - Can also depend on history, map to the reals, but unimportant for this talk.
- Goal: Given a collection of events $\mathcal{E}$, make $\mathcal{E}$-unbiased predictions: for each $E \in \mathcal{E}$:

$$\left\| \sum_{t=1}^{T} E(x_t, \hat{s}_t) \cdot (\hat{s}_t - s_t) \right\|_\infty \leq \alpha$$

- Calibration is the special case of $\{E(\hat{s}_t) = 1[\hat{s}_t = s]\}_{s \in C}$
  - Exponentially in d many events

Questions

- Calibration is useful but hard to obtain.. So:

1. Are predictions that are unbiased subject to only a modest (polynomial) number of conditioning events useful for decision making, and
2. Can we efficiently make $\mathcal{E}$-unbiased predictions for modestly sized collections $\mathcal{E}$?

Warmup: (Internal) Regret

- Consider a repeated interaction with action/state sequence:

$$\pi^T = \{(a_1, s_1), (a_2, s_2), \dots, (a_T, s_T)\}$$

- The agent has (external) Regret α if:

$$\sum_{t=1}^{T} u(a_t, s_t) \geq \min_{a \in A} \sum_{t=1}^{T} u(a, s_t) - \alpha$$

- The agent has internal regret α if for every $a \in A$:

$$\sum_{t=1}^{T} 1[a_t = a] \cdot u(a_t, s_t) \geq \min_{a' \in A} \sum_{t=1}^{T} 1[a_t = a] \cdot u(a', s_t) - \alpha$$

"On the subsequence on which the agent played a, a was a best response"

- If all agents in an interaction have α internal regret, play is a $k\alpha$-approximate correlated equilibrium.

Warmup: Internal Regret

- Best responding to calibrated forecasts yields no internal regret, but not necessary.
- Let $E_{u,a}(\hat{\mathrm{s}}) = 1[\mathrm{a} = BR(u, \hat{s})]$, $\mathcal{E}_u = \{E_{u,a}\}_{a \in A}$
  - The events that each action a is a best response to $\hat{s}$ for u.

Claim: If predictions $\hat{s}_1, \dots, \hat{s}_T$ have $\mathcal{E}_u$-bias α, then if an agent uses best response policy $f_u^*(\hat{s}) = BR(u, \hat{s})$, they have internal regret 2α.

- Only $k = |A|$ conditioning events, rather than 2^d
- If there are many agents with utility functions $u \in U$, can give unbiased predictions wrt $E_U = \bigcup_{u \in U} \mathcal{E}_u$ and obtain guarantees for all agents.

Warmup: Internal Regret

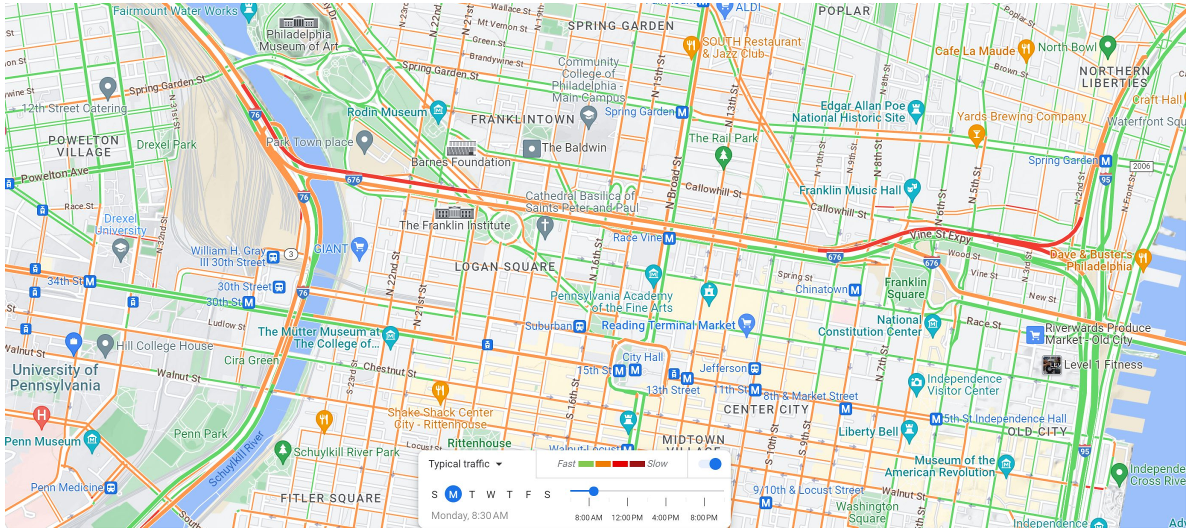
Claim: If predictions $\hat{s}_1, \ldots, \hat{s}_T$ have $\mathcal{E}_u$-bias α, then if an agent uses best response policy $f_u^*(\hat{s}) = BR(u, \hat{s})$, they have internal regret 2α.

Proof: Fix a pair of actions $a, a' \in A$.

$$\begin{aligned}\sum_t 1[a_t = a]u(a, s_t) &= \sum_t 1[BR(u, \hat{s}) = a]u(a, s_t) \\ &= u\left(a, \sum_t 1[BR(u, \hat{s}) = a]s_t\right) \\ &\geq u\left(a, \sum_t 1[BR(u, \hat{s}) = a]\hat{s}_t\right) - \alpha \\ &\geq u\left(a', \sum_t 1[BR(u, \hat{s}) = a]\hat{s}_t\right) - \alpha \\ &\geq \sum_t 1[a_t = a]u(a', s_t) - 2\alpha\end{aligned}$$

Generalizing: Online Combinatorial Optimization

- A set B of d *base* actions.
  - E.g. roads in a network
- A feasible set of actions $A \subseteq 2^B$
  - E.g. $s \to t$ paths in a network



- State s encodes gain $g_{u,e}(s)$ for each $s \in B$
- Utility for an action $a \in A$: $u(a, s) = \sum_{e \in a} g_{u,e}(s)$
- Up to 2^d actions, but d-dimensional linear structure.
- Expert learning the special case of $A = B$.

Generalizing: Subsequence Regret

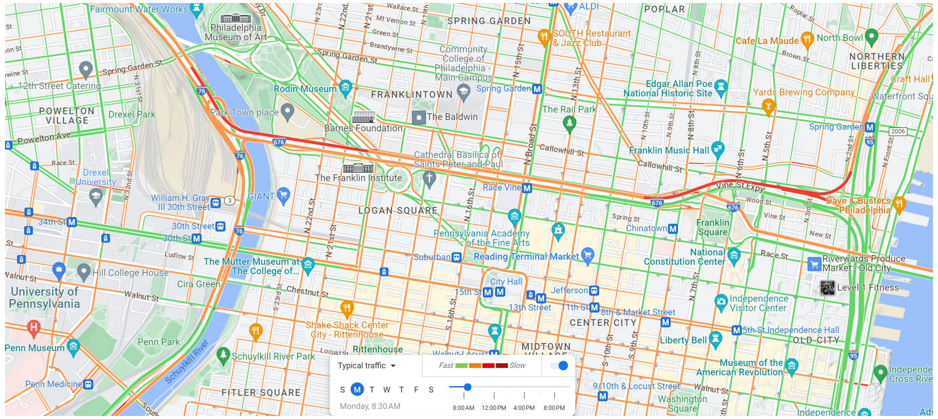
- Given a collection $\mathcal{E}$ of subsequence indicator functions $E(x_t, a_t)$, an agent has α-subsequence regret if for every $E \in \mathcal{E}$:
$$\sum_{t=1}^{T} E(x_t, a_t) u(a_t, s_t) \geq \min_{a' \in A} \sum_{t=1}^{T} E(x_t, a_t) \cdot u(a', s_t) - \alpha$$
  - Internal regret is the special case of events $E_a(a_t) = 1[a_t = a]$
- Let $E_{u,b}(\hat{s}_t) = 1[b \in BR(u, \hat{s}_t)]$, $\mathcal{E}_u = \{E_{u,b}\}_{b \in B}$
- Let $\mathcal{E} \times \mathcal{E}_u = \{E\left(x_t, BR(u, \hat{s}_t)\right) \cdot E_{u,b}(\hat{s}_t) \colon E \in \mathcal{E}, E_{u,b} \in \mathcal{E}_u\}$

Online Combinatorial Optimization

- Theorem: For any collection of events $\mathcal{E}$, if forecasts $\hat{s}_1, \ldots, \hat{s}_T$ have $(\mathcal{E} \times \mathcal{E}_u)$-bias at most α, then the best response policy has subsequence regret over $\mathcal{E}$ at most $2\alpha d$.

- Requires unbiasedness on only $d \cdot |\mathcal{E}|$ events.
- Can ask for this simultaniously for m agents with different u's and action sets, with $m \cdot d \cdot |\mathcal{E}|$ events…

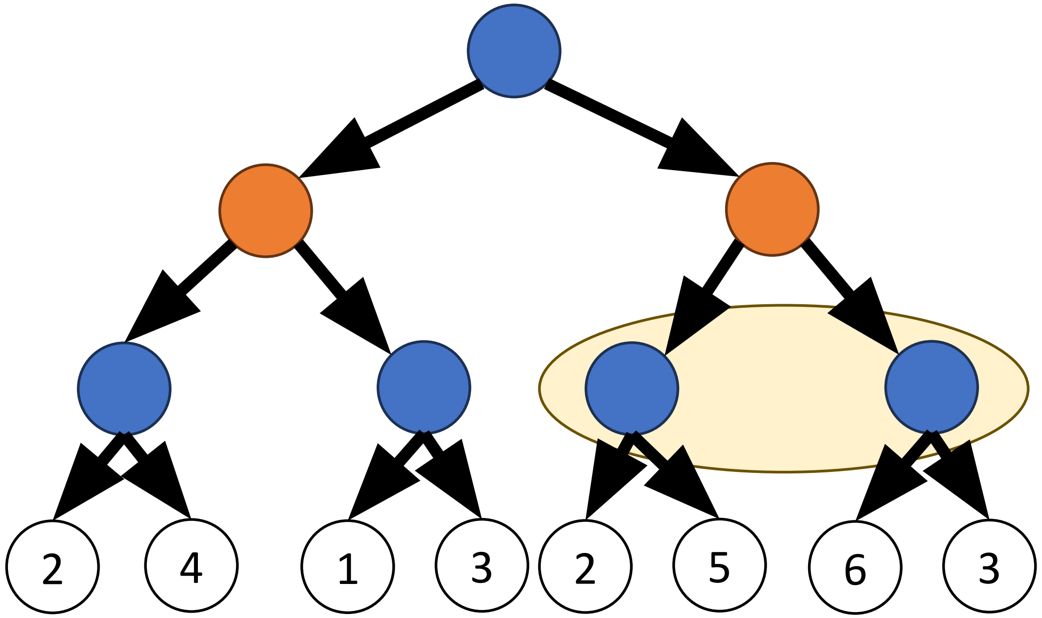
Online Combinatorial Optimization



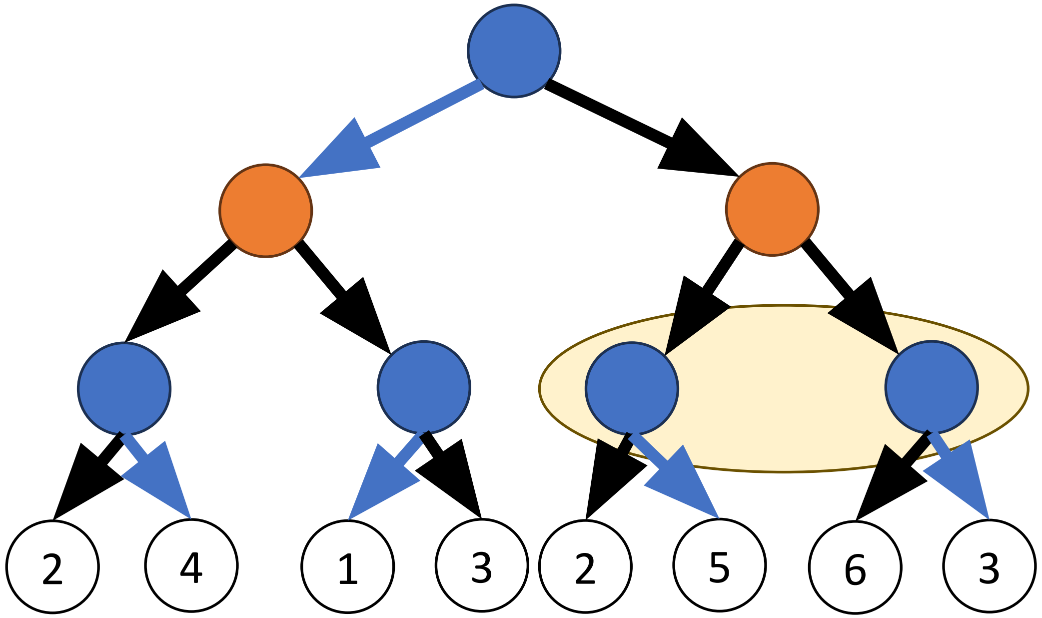
e.g. can publish traffic reports that are a good idea to follow (no regret) simultaniously for *every* agent who might have different source/destination pairs. Not just overall but also:

- On Rainy Days
- On Mondays
- On National Holidays
- On days when the best route involves I-76
- On days when the best route takes surface roads right after a Phillies game
- …

Extensive Form Games

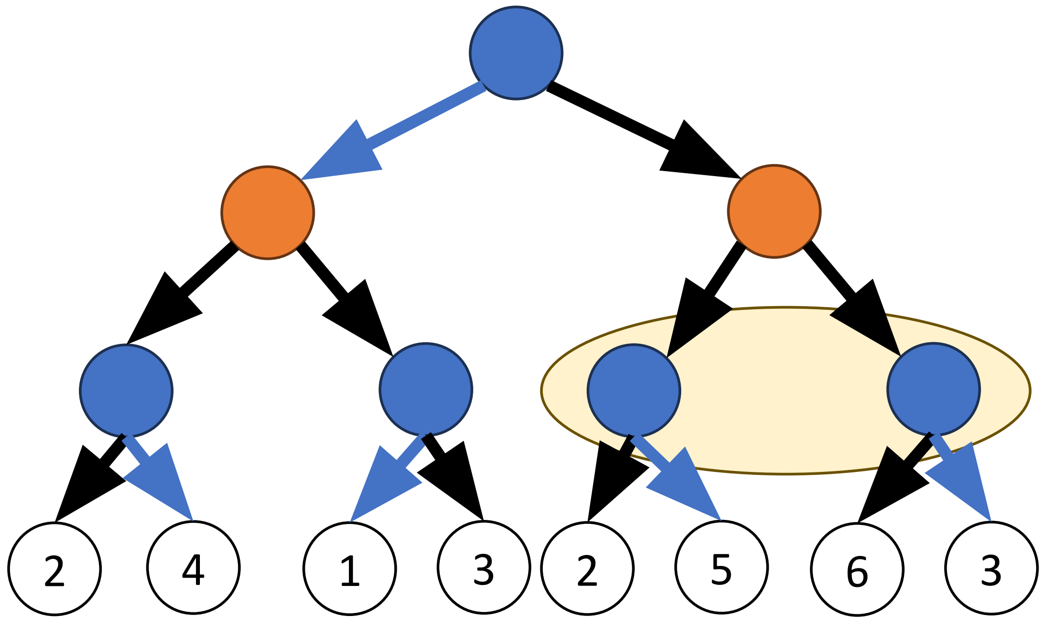


Extensive Form Games



Action space exponentially large in the game tree.

Extensive Form Games



A strategy makes a subset of the leaves *reachable*. Payoff corresponds to the leaf also made reachable by opponents.

Extensive Form Games

- Let $\mathcal{L}$ be the set of leaves in an extensive form game.
- Fix a player i. For $\ell \in \mathcal{L}$, let $r_i(\ell)$ be the reward for i at ℓ.
- Given a strategy a_i for player i, let $R(a_i) \in \{0,1\}^{|\mathcal{L}|}$ be the indicator vector for the set of leaves a_i makes reachable.
- The reward for agent i for playing a_i can be written as:

$$u(a_i, a_{-i}) = \langle R(a_i), g(a_{-i}) \rangle$$

Where $g(a_{-i}) \in \mathbb{R}^{|\mathcal{L}|}$ is $g_\ell(a_{-i}) = r_i(\ell) \cdot \Pr[\text{opponents make } \ell \text{ reachable}]$

- An instance of online combinatorial optimization…

Extensive Form Games

- So can efficiently obtain subsequence regret in extensive form games for any polynomial number of subsequences.
- By choosing the right collection of subsequences, we recover and generalize existing notions of extensive form regret:
  - Counterfactual regret
  - Regret to Causal deviations
  - …
- Converge to notions of extensive-form correlated equilibrium.

Efficiently Making $\mathcal{E}$-Unbiased Predictions
(Applying a Multiobjective Optimization Strategy from [LNPR22])

- A multiobjective optimization problem:
  - For each $E \in \mathcal{E}$, $i \in [d]$, $\sigma \in \{-1,1\}$ we want:
$$\sigma \cdot \sum_{t=1}^{T} E(x_t, \hat{s}_t) \cdot (\hat{s}_t - s_t)_i \le \alpha$$
  - Run a no regret algorithm (like Exponential Weights) over $2|\mathcal{E}|d$ experts (E, i, σ) with per-round gains defined as:
$$g_{E,i,\sigma}^t = \sigma E(x_t, \hat{s}_t)(\hat{s}_t - s_t)_i$$
  - No-regret guarantee is:
$$\sum_{t=1}^{T} \sum_{E,i,\sigma} w_{E,i,\sigma}^t \sigma E(x_t, \hat{s}_t)(\hat{s}_t - s_t)_i \ge \max_{E,i,\sigma} \sigma \cdot \sum_{t=1}^{T} E(x_t, \hat{s}_t) \cdot (\hat{s}_t - s_t)_i - O\left(\sqrt{T \cdot \log(|\mathcal{E}|d)}\right)$$
  - So if we can make predictions $\hat{s}_t$ such that $\mathbb{E}_{\hat{s}_t}\left[\sum_{E,i,\sigma} w_{E,i,\sigma}^t \sigma E(x_t, \hat{s}_t)(\hat{s}_t - s_t)_i\right] \le 0$ for all s_t…
  - We'll guarantee $\max_{E,i} \left|\sum_{t=1}^{T} E(x_t, \hat{s}_t) \cdot (\hat{s}_t - s_t)_i\right| \le O\left(\sqrt{T \cdot \log(|\mathcal{E}|d)}\right)$

Efficiently Making $\mathcal{E}$-Unbiased Predictions
(Applying a Multiobjective Optimization Strategy from [LNPR22])

- Goal: Find a distribution over predictions s.t. for all s_t:

$$\mathbb{E}_{\hat{s}_t}\left[\sum_{E,i,\sigma} w^t_{E,i,\sigma}\sigma E(x_t,\hat{s}_t)(\hat{s}_t - s_t)_i\right] \leq 0$$

- If the adversary first committed to a distribution over s_t, we could just play $\hat{s}_t = \mathbb{E}[s_t]$.
- By the minimax theorem, a solution exists --- we just need to find it!
- A minimax equilibrium computation…
- Difficulty: Exponentially large action spaces.

Efficiently Making $\mathcal{E}$-Unbiased Predictions

- We can compute a minimax strategy using learning dynamics. Repeated play in which:
  - The adversary uses a no-regret learning algorithm, and…
  - The learner best responds.
- The objective $\mathbb{E}_{\hat{s}_t}\left[\sum_{E,i,\sigma} w_{E,i,\sigma}^t \sigma E(x_t, \hat{s}_t)(\hat{s}_t - s_t)_i\right]$ is linear in the adversary's action s_t
  - So the adversary can efficiently play "Follow the Perturbed Leader"
- The objective is a complicated function of the learner's action $\hat{s}_t$… But to play so as to obtain value ≤ 0, it suffices to play $\hat{s}_t = \mathbb{E}[s_t]$.
  - i.e. can just "copy" the adversary's strategy.

Efficiently Making $\mathcal{E}$-Unbiased Predictions

Theorem: For any set of events $\mathcal{E}$ and any $\alpha > 0$, there is an online prediction algorithm that can make d-dimensional adversarial predictions over T rounds such that their worst-case $\mathcal{E}$-bias is at most α for:

$$\alpha \leq \sqrt{\log(d|\mathcal{E}|T) + T}$$

The per-round running time is polynomial in d and T.

- For disjoint events, running time decreases to polylog(T).
- More refined bounds (less bias for shorter subsequences)

Upshot

- Calibration incentivizes downstream agents to treat predictions as correct, but has exponentially growing complexity…
  - The complexity does not stem from the exponentially large action space, but the exponentially large number of conditioning events.
  - Can efficiently make predictions that are unbiased subject to polynomially many events.

Upshot

- For particular utility functions, polynomially many events are enough.
  - A useful algorithm design paradigm --- you only have one utility function!
- Can design “coordination mechanisms” for whole classes of utility functions
  - E.g. for all source-destination pairs in a routing game.
  - Predictions agents are incentivized to follow, and lead to no internal regret/correlated equilibrium.
  - Requires much less agent sophistication than running their own no-internal-regret algorithm.
- Concrete Mechanism Design Application
  - Sequential principal agent problems without a prior, a la Camara, Hartline, and Johnson with an exponentially improved dependence on the state space. (joint work with Natalie Collina and Han Shao)

Upshot

- Other applications…
  - Uncertainty Quantification:
    - In multiclass classification problems can produce class scores that can be treated as probabilities for producing prediction sets of different coverage probabilities.
    - In regression problems can produce functions that can be treated as label CDFs to produce prediction intervals of different coverage probabilities.

Thanks!