Meritocratically Fair Algorithms for Infinite and Contextual Bandits

M. Joseph, M. Kearns, J. Morgenstern, S. Neel, A. Roth

University of Pennsylvania

High-Level Motivation

- Machine learning can be *unfair* in many ways: biased data; different populations with different properties; less data about minorities, etc.
- How do we define *fair learning*? What is the performance cost of fairness?

Previous Work

- JKMR16 [1] studied fairness for finite contextual bandits
- **Problem 1**: unrealistic assumptions (one individual per group per day; choose exactly one individual per day – artificial inter-group competition?)
- **Problem 2**: results do not scale well when number of arms is large

Finite Setting

- Goal: address Problem 1 above
- In each round $\mathbf{t}$ we see set $\mathbf{C_t}$ of at most $\mathbf{k}$ contexts in $\mathbb{R}^{\mathbf{d}}$, choose a subset $\mathbf{P_t} \subset \mathbf{C_t}$ of exactly $\mathbf{m}$ contexts, and observe noisy linear reward $\mathbf{r_i^t} = \langle \beta, \mathbf{x_i^t} \rangle + \epsilon_i^t$ for each $\mathbf{i} \in \mathbf{P_t}$
- Addresses problem 1: can see multiple individuals per population per round, can choose multiple individuals per round
- Group membership can be encoded in context in $\mathbb{R}^{\mathbf{d}}$ or not
- Goal: maximize $\sum_t \sum_{i \in P_t} \mathbb{E}[r_i^t]$, measure performance by regret $\mathbf{R(T)} = \sum_t [\mathbb{E}[\sum_{i \in P_t^*} r_i^t] - \mathbb{E}[\sum_{j \in P_t} r_j^t]]$ (loss from choosing subset $\mathbf{P_t}$ instead of best expected subset $\mathbf{P_t^*}$ across rounds)
- Models a program that learns to grant loans by granting $\mathbf{m}$ loans daily

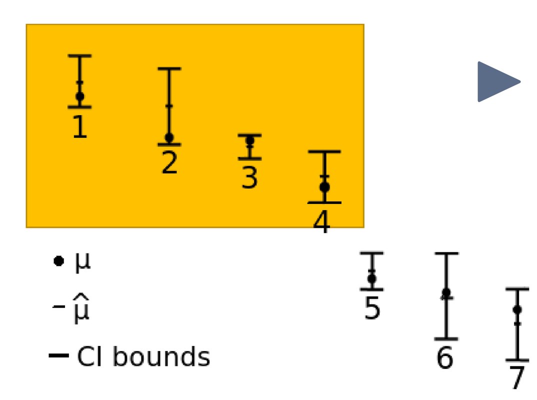
General Fairness Definition

- Algorithm $\mathcal{A}$ is **fair** if (whp) for all $\mathbf{t} \in \mathbf{T}$ and for all $\mathbf{i}, \mathbf{j} \in \mathbf{C_t}$ $\mathbb{E}[r_i^t] \geq \mathbb{E}[r_j^t] \Rightarrow \pi_i^t \geq \pi_j^t$ where $\pi_i^t = \mathbb{P}[\text{choose } \mathbf{i} \text{ in round } \mathbf{t}]$ (omitting histories for simple notation)

Interpreting Fairness Definition

- "Whp, never more likely to choose a worse arm than a better arm"
- Optimal policies always play the expected best arm and therefore are fair. Challenge: how to *learn* the optimal policy fairly?

A Fair Algorithm: RidgeFair$_\mathbf{m}$



RidgeFair$_5$ would select arms 1, 2, 3, and 4, and also select one of 5, 6, and 7 at random.

- Uses confidence intervals around estimated means to reason about relative quality; fairness forces *chaining*
- In round $\mathbf{t}$: Choose all arms in highest connected component of confidence intervals, then choose the last arms by randomizing when you reach a connected component in which you cannot choose all arms

FairUCB [1] vs RidgeFair$_\mathbf{m}$

- Encoded in our setting (using contexts in $\mathbb{R}^{\mathbf{dk}}$) FairUCB achieves regret $\mathbf{R(T)} = \mathbf{O(max[T^{4/5}k^{6/5}d^{3/5}, k^3])}$
- RidgeFair$_\mathbf{m}$ achieves regret $\mathbf{R(T)} = \mathbf{O(dk^2\sqrt{T})}$
- Improvement via better (and more technical) confidence intervals for $\hat{\beta}$
  - Uses martingale matrix concentration results from APS11 [2]

Infinite Setting

- Goal: address Problem 2 above
- In each round $\mathbf{t}$ we see a convex set $\mathbf{C_t}$ of choices contained in a ball of radius $\mathbf{r}$, select exactly one, and observe (single) noisy reward $\mathbf{r_t} = \langle \beta, \mathbf{x} \rangle + \epsilon_t$
- Goal: maximize $\sum_t \mathbb{E}[r_t]$, measure performance by regret $\mathbf{R(T)} = \sum_t \mathbb{E}[r_t^* - r_t]$ where $\mathbf{r_t*}$ is an optimal choice in round $\mathbf{t}$ and $\mathbf{r_t}$ is the actual choice

A Fair Algorithm: FairGap

- Uses convexity of each $\mathbf{C_t}$: optimal point must be an *extremal* point
- Plays randomly until confidence interval around $\hat{\beta}$ shrinks enough to separate optimal extremal point from suboptimal extremal points
- Performance thus depends on $\mathbf{\Delta_{gap}}$ – the "gap" in expected reward between an optimal and next sub-optimal extremal point
- Instance-dependent regret bound:
$$\mathbf{R(T)} = \tilde{\mathbf{O}}\left(\frac{\mathbf{r^6R^2}}{\kappa^2\lambda^2\mathbf{\Delta^2_{gap}}}\right)$$
where $\kappa = \mathbf{1} - \mathbf{r}\sqrt{\frac{\mathbf{2}}{\mathbf{T}\lambda}}$ and $\lambda = \mathbf{min_{1 \leq t \leq T}} \left[\lambda_{\mathbf{min}}(\mathbb{E}_{\mathbf{x_t} \sim \mathbf{C_t}}[\mathbf{x_t}^{\mathbf{T}}\mathbf{x_t}])\right]$
- Regret independent of $\mathbf{k}$

Instance-dependent Lower Bound

- **Thm:** Let $\mathbf{C_t} = [-1, 1]^{\mathbf{d}}$ for each $\mathbf{t}$ and choose some $\beta \in [-1, 1]^{\mathbf{d}}$. Then for every $\mathbf{\Delta_{gap}}$ there exists an instance distribution for which any fair algorithm whp experiences $\tilde{\mathbf{\Omega}}(\mathbf{1}/\mathbf{\Delta_{gap}})$ regret.
  - Adapts Bayesian lower-bound argument from JKMR16 [1]
- (Some) instance-dependence is therefore necessary for any fair algorithm in the infinite setting
  - FairGap's $\mathbf{O}(\mathbf{1}/\mathbf{\Delta^2_{gap}})$ regret is almost tight

Instance-independent Lower Bound

- **Thm:** Let $\mathbf{C_t}$ be $\mathbf{S^1}$ (the unit circle) for each $\mathbf{t}$. Then for any $\beta \in \mathbf{S^1}$, no fair algorithm achieves nontrivial regret.
- Consequence of $\mathbf{\Delta_{gap}} = \mathbf{0}$ – continuity means you can never actually identify an optimal point

References

[1] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth.
Fairness in learning: Classic and contextual bandits.
In *NIPS 2016.*

[2] Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári.
Improved algorithms for linear stochastic bandits.
In *NIPS 2011.*