Fairness in Learning: Classic and Contextual Bandits

M. Joseph, M. Kearns, J. Morgenstern, A. Roth
{majos, mkearns, jamiemor, aaroth}@cis.upenn.edu

Penn
UNIVERSITY of PENNSYLVANIA

High-Level Motivation

- Machine learning can be unfair in many ways: data that encodes existing biases; data collection feedback loops; different populations having different properties; less data about minority populations ...
- How do we define "fair learning"?
- What is the performance cost of being fair?

General Problem Setting

- We study the *bandits* setting: $\mathbf{k}$ arms, on day $\mathbf{t} \in \mathbf{T}$ choose arm $\mathbf{i^t}$ and observe noisy reward $\mathbf{r^t_{i^t}}$
- Goal: maximize $\sum_t \mathbb{E}[\mathbf{r^t_{i^t}}]$, measure performance by regret $\mathbf{R(T)} = \sum_t [\max_{i \in [k]} \mathbb{E}[\mathbf{r^t_i}] - \mathbf{r^t_{i^t}}]$
  - Models a program that learns to grant loans to $\mathbf{k}$ different groups by granting loans to one member of one group each day

General Fairness Definition

- Algorithm $\mathcal{A}$ is **fair** if with probability $\geq 1 - \delta$, for all days $\mathbf{t} \in \mathbf{T}$ and for all $\mathbf{i}, \mathbf{j} \in [\mathbf{k}]$

$$\mathbb{E}[\mathbf{r^t_i}] \geq \mathbb{E}[\mathbf{r^t_j}] \Rightarrow \pi^t_{i|h_1,\ldots,h_{t-1}} \geq \pi^t_{j|h_1,\ldots,h_{t-1}}$$

  where $\pi^t_{i|h_1,\ldots,h_{t-1}} =$ $\mathbb{P}[\text{choose } \mathbf{i} \text{ in round } \mathbf{t} \text{ after observing } \mathbf{h_1, \ldots, h_{t-1}}]$.
- "With high probability, never more likely to choose a worse arm than a better arm"

Why is Fairness Hard?

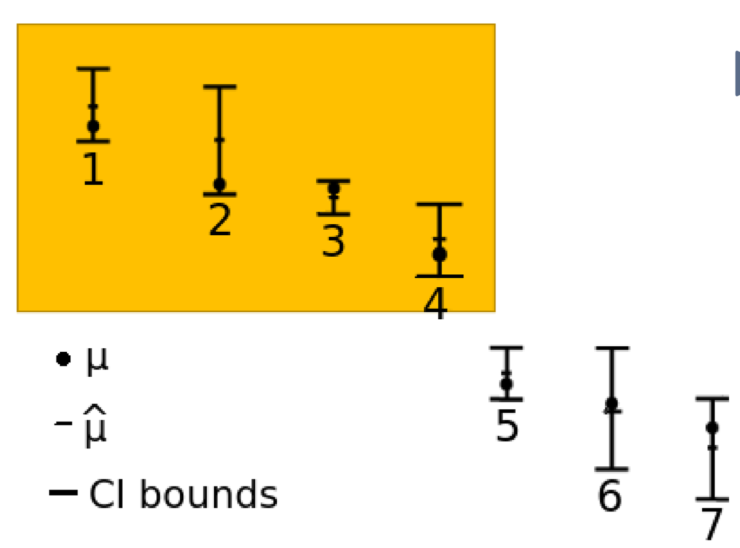
- Optimal policies always play the expected best arm and therefore are fair. Challenge: how to *learn* the optimal policy fairly?

Classic Bandits Setting

- μ_i for each arm $\mathbf{i}$ such that for all $\mathbf{i}$ and $\mathbf{t}$ $\mathbb{E}[\mathbf{r^t_i}] = \mu_i$
- Fair: $\mu_i \geq \mu_j \Rightarrow \pi^t_{i|h_1,\ldots,h_{t-1}} \geq \pi^t_{j|h_1,\ldots,h_{t-1}}$
- "With high probability, never more likely to choose an arm with lower μ than an arm with higher μ"

A Fair Classic Bandit Algorithm: FairBandits

- Uses confidence intervals around estimated means to reason about relative quality; fairness forces *chaining*



FairBandits plays randomly from chain (Arms 1 to 4)

- In round $\mathbf{t}$: pick uniformly at random from "chain" of top arms (top connected component of overlapping confidence intervals)

Cost of Fairness in Classic Bandits

- FairBandits regret upper bound $\mathbf{R(T)} = \tilde{\mathbf{O}}(\sqrt{\mathbf{k^3 T}})$
- Regret lower bound (any fair algorithm) $\mathbf{R(T)} = \Omega(\mathbf{k^3})$, while $\mathbf{R(T)} = \tilde{\Theta}(\sqrt{\mathbf{kT}})$ (unfair)

Contextual Bandits Setting

- Function $\mathbf{f_i} \in \mathbf{C}$ for $\mathbf{i} \in [\mathbf{k}]$; $\mathbf{x^t_i} \in \mathbb{R}^d$ for $\mathbf{t} \in \mathbf{T}$, $\mathbf{i} \in [\mathbf{k}]$ such that $\mathbb{E}[\mathbf{r^t_i}] = \mathbf{f_i(x^t_i)}$
- Fair: $\mathbf{f_i(x^t_i)} \geq \mathbf{f_j(x^t_j)} \Rightarrow \pi^t_{i|h_1,\ldots,h_{t-1}} \geq \pi^t_{j|h_1,\ldots,h_{t-1}}$
- "With high probability, never more likely to choose an arm with lower $\mathbf{f(x^t)}$ than an arm with higher $\mathbf{f(x^t)}$"

Fair Contextual Bandits and KWIK Learning

- $\mathcal{C}$ is KWIK-learnable [1] with poly KWIK bound $\Leftrightarrow$ $\mathcal{C}$ can be learned fairly with poly regret
- For $\mathbf{d}$-dimensional *linear functions*, KWIK bounds [2] imply fair learning with $\mathbf{R(T)} = \tilde{\mathbf{O}}\left(\max\left(\mathbf{T^{4/5} k^{6/5} d^{3/5}}, \mathbf{k^3}\right)\right)$
- For $\mathbf{d}$-dimensional *conjunctions*, KWIK bounds [3] imply that no fair learning algorithm has a worst-case regret bound better than $\mathbf{R(T)} = \Omega(\mathbf{2^d})$

References

[1] Lihong Li, Michael L Littman, Thomas J Walsh, and Alexander L Strehl. Knows what it knows: a framework for self-aware learning. *Machine learning*, 82(3):399–443, 2011.

[2] Alexander L Strehl and Michael L Littman. Online linear regression and its application to model-based reinforcement learning. In *Advances in Neural Information Processing Systems*, pages 1417–1424, 2008.

[3] Lihong Li. *A unifying framework for computational reinforcement learning theory*. PhD thesis, Rutgers, The State University of New Jersey, 2009.