Accounting for **fairness** complicates **online learning for decision making.** When is this hard and when is it easy?

Constrained, Multi-objective Optimization with Contextual Multi-armed Bandits

Henry Zhu, Alex Chohlas-Wood, Madison Coots, Sharad Goel, Emma Brunskill

INTRO

- **Many definitions of fairness** may be important in practice (e.g. in outcomes or spending)
- We study fairness in decision making in the **online setting** and ask:
  - When are **optimistic** algorithms **consistent** (sublinear regret) and **computationally efficient**?

SETUP

- Policy π, contexts X, actions A.
- Rewards r_{XA} are unknown.
- Utility function $U(\pi, r)$
- Valid policy set Π
- Goal:

$$\max_{\pi} U(\pi, r) \text{ s.t. } \pi \in \Pi$$

- Eg. of utility function including fairness of outcomes

$$U(\pi, r) = \mathbb{E}_{\pi}[r_{XA}] + F(\pi, r),$$
$$F(\pi, r) = \left| \mathbb{E}_{\pi}\left[r_{XA} \,\middle|\, X \in G_1\right] - \mathbb{E}_{\pi}\left[r_{XA} \,\middle|\, X \in G_2\right] \right|$$

RESULTS

- **Utility optimism has sublinear regret** under very mild conditions, whereas **point-wise optimism can incur linear regret in general.**
- However, **point-wise optimism is computationally efficient** whereas **utility optimism is not.**
  - Under convex objective + convex constraints, point-wise -> convex optimization but utility optimism -> non-convex optimization.
- Point-wise optimism has sublinear regret under when the objective is **monotonic.**
  - Under these conditions, point-wise and utility optimism are equivalent.
- Validate results across 3 domains with over 10 different objectives and constraints.

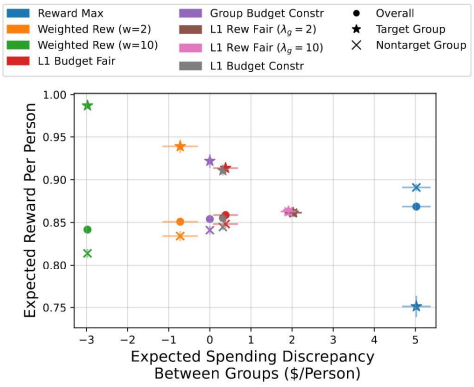
Experimental Domains

I. **Increasing court appearances** under budgets with equity in spending + outcomes. (Figures 1, 2)

II. **Loan approval** with actionable recourse.

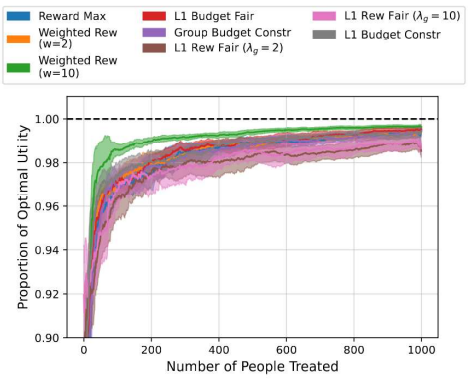
III. **Optimal drug dosing** with budgets and equity in spending.

Figure 1. Court Appearances: Reward v. spending discrepancies



On the left is a scatter plot with different colored and shaped dots. The colors correspond to different policies learned under different objective function and shapes correspond to different racial groups. The takeaway is that changing the objective function meaningfully changes the policy learned.

Figure 2. Court Appearance: Sublinear regret



On the right is a line plot with lines of different colors. The colors correspond to different policies learned under different objective functions. The takeaway is that for all the different objectives considered, the algorithm is able to learn the corresponding optimal policy quickly.