Private-Fair-Greedy: A DP Algorithm for Group Meritocratic Fairness in Contextual Linear Bandits

Esther An Andrew Palacci Phevos Paschalidis

Motivation & Contributions

Every year, an employer receives K applications for a single position. After hiring their preferred candidate, the employer observes a scalar reward reflecting job performance. The employer's goal is to maximize the total reward over T years, in part by using past experience to inform their estimate of future candidate performance. We are interested in finding a private and fair algorithm for this problem.

We model the setting as a contextual linear bandit problem and present `Private-Fair-Greedy`, which satisfies joint differential privacy and group meritocratic fairness: assuming each candidate belongs to a sensitive group, the policy learns to select the candidate with the best performance *relative to the distribution of others from its sensitive group*.

Problem Formulation

Consider the contextual linear bandit problem with K arms, each corresponding to a sensitive group. At each time step $t \in [T]$, the agent is presented with K feature vectors $D_t := \{X_{k,t}\}_{k\in[K]}$, each with associated reward $y_{k,t} := \langle X_{k_t,t}, \theta^* \rangle + \eta_t$, where $\mathbb{X}_{k,t} \in \mathbb{R}^d$, θ^* is an unknown preference vector and $\eta_{k,t}$ sub-Gaussian and zero-meaned. The agent selects action $k_t \in [K]$ and observes the noisy reward $y_{k_t,t}$.

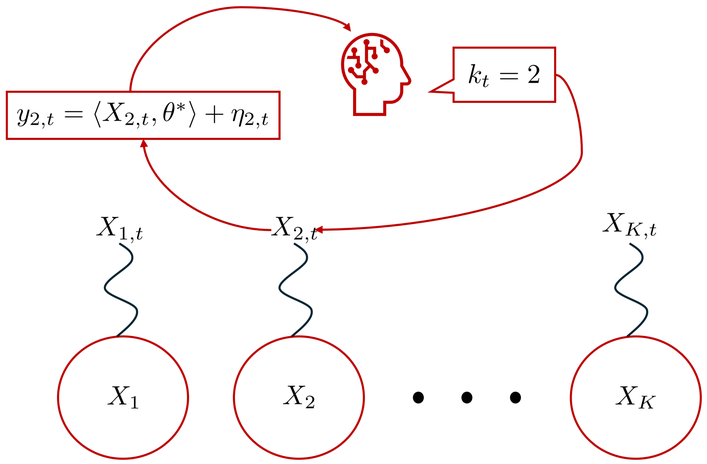

Figure 1. At each time step t, an agent selects an action k_t and observes a noisy reward.

Definitions

Relative Rank. The relative rank of a candidate x is $\mathcal{F}_k(\langle x, \theta^* \rangle)$, or the probability that another element from its sensitive group results in lower reward.

Group Meritocratic Fairness: We assume that, for each $k \in [K]$, each candidate $X_{k,t}$ is drawn independently from the same unknown distribution X_k. Then, $\langle X_k, \theta^* \rangle$ is the reward distribution for group k with CDF $\mathcal{F}_k(y) = \mathbb{P}(\langle X_k, \theta^* \rangle \le y)$. A policy $\{k_t\}_{t=1}^T$ is group meritocratic fair (GMF) if, for all $t \in [T]$ and $k \in [K]$,

$$\mathcal{F}_{k_t}(\langle X_{k_t,t}, \theta^* \rangle) \ge \mathcal{F}_k(\langle X_{k,t}, \theta^* \rangle).$$

Fair Pseudo-Regret: Let $T \in \mathbb{N}$ be the time horizon, $\{k_t\}_{t=1}^T$ the chosen policy, and $\{k_t^*\}_{t=1}^T$ a GMF policy. Then, the cumulative fair pseudo-regret is

$$R(T) := \sum_{t=1}^{T} \left(\mathcal{F}_{k_t^*}(\langle X_{k_t^*,t}, \theta^* \rangle) - \mathcal{F}_{k_t}(\langle X_{k_t,t}, \theta^* \rangle) \right).$$

Joint Differential Privacy: Let $z_t = \{y_{k,t}\}_{k\in[K]}$. We say $\mathcal{T} = \{\mathcal{D}_t, z_t\}_{t\in[T]}$ and $\mathcal{T}' = \{\mathcal{D}'_t, z'_t\}_{t\in[T]}$ are t-neighbors if it holds that $(\mathcal{D}_{t'}, z_{t'}) = (\mathcal{D}'_{t'}, z'_{t'})$ for all $t' \neq t$ and $(\mathcal{D}_t, z_t)$ differ only in a single tuple $(X_{k,t}, y_{k,t})$ for some specific $k \in [K]$. Then, a randomized algorithm A is (ε, δ)-jointly differentially private if, for any t, any t-neighboring histories $\mathcal{T}$ and $\mathcal{T}'$, and any set of $T-1$ actions $B \in [K]^{T-1}$, it holds that

$$\mathbb{P}(A(\mathcal{T})_{-t} \in B) \le e^{\varepsilon} \mathbb{P}(A(\mathcal{T}')_{-t} \in B) + \delta.$$

The Private-Fair-Greedy Policy

The goal of the fair contextual linear bandit setting is to design algorithms that learn to be GMF by minimizing fair pseudo-regret.

The Algorithm

At each time step, `Private-Fair-Greedy` utilizes past data to compute a private regression estimate of the feature vector, which it uses to privately estimate the relative rank of available candidates. It acts greedily, selecting the candidate with highest relative rank. `Private-Fair-Greedy` thus extends `Fair-Greedy` [6], leveraging differentially private techniques for regression and rank estimation.

We manage the privacy budget by dividing (ε, δ) between the regression and relative rank estimation tasks. Our private subroutines each allocate their respective budgets across T time steps. To justify this two-step, independent division, we rely on a continual composition theorem that shows that interactive continual tasks share the privacy guarantees of non-interactive ones [7].

Continually Private Regression

A non-private closed form solution for linear ridge regression is

$$\hat{\theta}_{\tilde{t}}^{\text{ridge}} = (X_{1:\tilde{t}}^\top X_{1:\tilde{t}} + \lambda \mathbb{I}_d)^{-1} X_{1:\tilde{t}}^\top y_{1:\tilde{t}},$$

where λ is a regularization parameter. A private version of the regression thus adds sufficient noise to $X_{1:\tilde{t}}^\top X_{1:\tilde{t}}$ and $X_{1:\tilde{t}}^\top y_{1:\tilde{t}}$. We define $A := \begin{bmatrix} X_{1:T} & y_{1:T} \end{bmatrix} \in \mathbb{R}^{T\times(d+1)}$, with $A_{\tilde{t}}$ holding the top $\tilde{t}$ rows of A. Let $M_{\tilde{t}} = A_{\tilde{t}}^\top A_{\tilde{t}}$. Then, $X_{1:\tilde{t}}^\top X_{1:\tilde{t}}$ is the top left $d \times d$ submatrix of $M_{\tilde{t}}$ and $X_{1:\tilde{t}}^\top y_{1:\tilde{t}}$ is the first d entries of its last column. To perform a private regression, we therefore need to maintain a noisy estimate for only $M_{\tilde{t}}$.

$M_{\tilde{t}+1} = M_{\tilde{t}} + \begin{bmatrix} X_{k_{\tilde{t}},\tilde{t}}^\top & y_{\tilde{t}} \end{bmatrix}^\top \begin{bmatrix} X_{k_{\tilde{t}},\tilde{t}}^\top & y_{\tilde{t}} \end{bmatrix}$, so $M_{\tilde{t}}$ is an incrementally updated sum of the data at individual time steps. We therefore use techniques from the framework of continual differential privacy to minimize noise addition. Specifically, we use a tree-based mechanism with the i^{th} leaf of the tree storing $\begin{bmatrix} X_{k_i,i}^\top & y_i \end{bmatrix}^\top \begin{bmatrix} X_{k_i,i}^\top & y_i \end{bmatrix}$ [3, 4]. Each node maintains a noisy sum of the data stored in its subtree. Since any cumulative sum can be obtained by combining at most $m = 1 + \lceil \log(T/2) \rceil$ intermediate values, we are able to achieve $O(\varepsilon_{\text{reg}}/\sqrt{\log T})$ noise for each $M_{\tilde{t}}$ rather than the $O(\varepsilon_{\text{reg}}/\sqrt{T})$ factor offered by standard (non-continual) composition theorems.

Private Relative Rank Estimation

The non-private relative rank estimate from [6], $\hat{F}_{k,t}(\langle X_{k,t}, \hat{\theta}_{\tilde{t}} \rangle)$, calculates the mean of indicator values $\mathbb{1}\{\langle X_{k',s}, \hat{\theta}_{\tilde{t}} \rangle \le \langle X_{k,t}, \hat{\theta}_{\tilde{t}} \rangle\}$ over a history of length N_t, performing this calculation for each arm $k \in [K]$ at every round $t \in [T]$. As this is a count query of 1's divided by the known history length, we use the Gaussian mechanism for zero-concentrated differential privacy to privatize this mean estimation at each time step.

To determine privacy parameter ρ_0 for each ρ_0-zCDP rank estimation release so that the total budget $(\varepsilon_{\text{rank}}, \delta_{\text{rank}})$ is respected across all releases, we employ composition for zCDP [2]. As indicators are in $\{0, 1\}$, the sensitivity of computing the mean over N_t items is $\Delta q_t = 1/N_t$. We thus use the Gaussian noise parameter $\sigma_t = \sqrt{T/2N_t^2\rho_{\text{rank}}}$, where

$$\rho_{\text{rank}} = \left(\sqrt{\log(1/\delta_{\text{rank}}) + \varepsilon_{\text{rank}}} - \sqrt{\log(1/\delta_{\text{rank}})} \right)^2.$$

References

[1] Yasin Abbasi-yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 24. Curran Associates, Inc., 2011.

[2] Mark Bun and Thomas Steinke. Concentrated differential privacy: Simplifications, extensions, and lower bounds, 2016.

[3] T-H Hubert Chan, Elaine Shi, and Dawn Song. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

[4] Cynthia Dwork, Moni Naor, Toniann Pitassi, and Guy N. Rothblum. Differential privacy under continual observation. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 715–724, New York, NY, USA, 2010. Association for Computing Machinery.

[5] Cynthia Dwork, Guy N Rothblum, and Salil Vadhan. Boosting and differential privacy. *Proceedings of the 51st Annual IEEE Symposium on Foundations of Computer Science*, (FOCS '10):51–60, 2010.

[6] Riccardo Grazzi, Arya Akhavan, John IF Falk, Leonardo Cella, and Massimiliano Pontil. Group meritocratic fairness in linear contextual bandits. *Advances in Neural Information Processing Systems*, 35:24392–24404, 2022.

[7] Monika Henzinger, Roodabeh Safavi, and Salil Vadhan. Concurrent composition for differentially private continual mechanisms, 2025.

[8] Michael Kearns, Mallesh M Pai, Ryan Rogers, Aaron Roth, and Jonathan Ullman. Robust mediators in large games. *arXiv preprint arXiv:1512.02698*, 2015.

Private-Fair-Greedy Satisfies Joint Differential Privacy

Theorem. Given a contextual linear bandit problem with time horizon $T \in \mathbb{N}$ and feature vectors and rewards bounded such that $\|X_{k,t}\| + \|y_t\| \le \tilde{L}$ for all $k \in [K]$ and $t \in [T]$, the `Private-Fair-Greedy` algorithm satisfies (ε, δ)-joint differential privacy.

Empirical Results

We evaluate `Private-Fair-Greedy` against non-private `Fair-Greedy` and the standard "Optimism in the face of uncertainty linear bandit algorithm" `OFUL` [1] on the U.S. Census Adult dataset, accessed via `folktables`. `Fair-Greedy` optimizes for GMF; `OFUL` maximizes expected reward subject to confidence bounds.

We model a sequential hiring scenario where an agent selects a candidate from a pool of $K = 4$ individuals from different racial groups in each round. The context vectors are constructed from the individuals' demographic features. The true reward is based on the chosen candidate's potential income, modeled as a linear function of their context vector. Ground-truth coefficients for the linear function are learned via regression on a hold-out training set.

The algorithms are evaluated over $T = 50{,}000$ rounds. `Private-Fair-Greedy` is run with privacy parameters $\varepsilon = 15$ and $\delta = 0.1$, with the budget allocated evenly across the regression and rank estimation tasks.

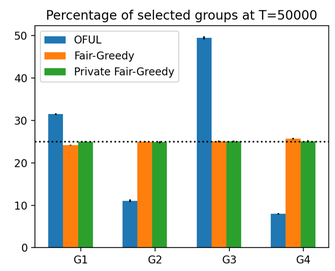

Figure 2. Percentage of times an arm from each sensitive group was chosen.

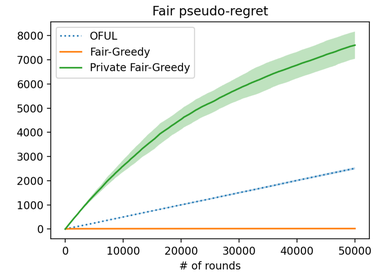

Figure 3. Cumulative fair pseudo-regret for each algorithm over $T = 50{,}000$ rounds.

Fairness Across Groups.

In Figure 2, the dotted line at 25% represents the expected selection rate per group under perfect demographic parity given $K = 4$. As anticipated, `Fair-Greedy` and `Private-Fair-Greedy` closely approximate parity across groups. `OFUL`, which does not incorporate fairness constraints, frequently selects arms from one group while neglecting others.

Fair Pseudo-Regret.

`Fair-Greedy` accrues near-zero cumulative fair pseudo-regret over the horizon, indicating that it closely approximates an optimal GMF policy (Figure 3). `OFUL` shows approximately linear growth in fair pseudo-regret: without fairness constraints, we expect consistent violations of GMF over time. `Private-Fair-Greedy` exhibits sublinear growth in cumulative fair pseudo-regret; however, the fair pseudo-regret is *much* higher than that of `OFUL`, suggesting that the noise introduced causes serious degradation even with privacy loss parameters $\varepsilon = 15, \delta = 0.1$.

Conclusion & Future Work

We extend the `Fair-Greedy` algorithm for the contextual linear bandit problem to construct `Private-Fair-Greedy`, which satisfies (ε, δ)-JDP. We also empirically evaluate the fairness and performance of `Private-Fair-Greedy` relative to `Fair-Greedy` and `OFUL`. In future work, we may extend our theoretical results to provide regret analysis for `Private-Fair-Greedy`. While we provide one method of privatizing `Fair-Greedy`, we recognize there exist additional approaches to implementing continually private regression and relative rank estimation. The contextual linear bandit setting remains relevant to high-stakes fields involving sensitive decisions about individuals. We hope developing private and fair algorithms will continue to enable safe and effective machine learning.