Fair Algorithms for Infinite Contextual Bandits *

Matthew Joseph, Michael Kearns, Jamie Morgenstern, Seth Neel, Aaron Roth
University of Pennsylvania

Abstract

We study fairness in infinite linear bandit problems. Starting from the notion of meritocratic fairness introduced in Joseph et al. [9], we expand their notion of fairness for infinite action spaces and provide an algorithm that obtains a sublinear but instance-dependent regret guarantee. We then show that this instance dependence is a necessary cost of our fairness definition with a matching lower bound. This provides a strong contrast with the traditional non-fair setting, where instance-independent regret bounds are achievable. Finally, we exhibit an action space in which fair algorithms cannot even obtain nontrivial instance-dependent bounds.

1 Introduction

The problem of repeatedly making choices and learning from choice feedback arises in a variety of settings, including granting loans, serving ads, and hiring. Encoding these problems in a *bandit* setting enables one to take advantage of a rich body of existing bandit algorithms. UCB-style algorithms, for example, are guaranteed to yield no-regret policies for these problems.

Joseph et al. [9], however, raises the concern that these no-regret policies may be *unfair*: in some rounds, they will choose options with lower expected rewards over options with higher expected rewards, for example choosing less qualified job applicants over more qualified ones. Joseph et al. [9] remedy this with no-regret algorithms which minimize mistreatment and are fair in the following sense: their algorithms (with high probability) never at any round place higher selection probability on a less qualified applicant than on a more qualified applicant. Our companion paper generalizes the problem setting and guarantees obtained, but still obtains results that scale polynomially in k, the number of bandits. This may be undesirable for large k, thus motivating the investigation of fair algorithms for the *infinite* bandit setting (the online linear optimization with bandit feedback problem [6]).

In Section 3 we provide such an algorithm. We then prove, subject to certain assumptions, a regret upper bound that depends on Δ_{gap}, an instance-dependent parameter based on the distance between the best and second-best extreme points in a given choice set. In Section 4 we show that this instance dependence is almost tight by exhibiting an infinite choice set satisfying our assumptions for which *any* fair algorithm must incur regret dependent polynomially on Δ_{gap}, separating this setting from the online linear optimization setting absent a fairness constraint. Finally, we justify our assumptions on the choice set by in Section 5 exhibiting a choice set that both violates our assumptions and admits *no* fair algorithm with nontrivial regret guarantees.

1.1 Related Work and Discussion of Our Fairness Definition

Fairness in machine learning has seen substantial recent growth as a subject of study, and many different definitions of fairness exist. We provide a brief overview here; see e.g. Berk et al. [1] and Corbett-Davies et al. [3] for detailed descriptions and comparisons of these definitions.

Many extant fairness notions are predicated on the existence of *groups*, and aim to guarantee that certain groups are not unequally favored or mistreated. In this vein, Hardt et al. [8] introduced the notion of *equality of opportunity*, which requires that a classifier's predicted outcome should be independent of a protected attribute (such as race) conditioned on the true outcome, and they and Woodworth et al. [11] have studied the feasibility and possible relaxations thereof. Similarly, Zafar et al. [12] analyzed an equivalent concurrent notion of (un)fairness they call *disparate mistreatment*. Separately, Kleinberg et al. [10] and Chouldechova [2] showed that different notions of group fairness may (and sometimes must) conflict with one another.

This paper, like Joseph et al. [9], departs from the work above in a number of ways. We attempt to capture a particular notion of *individual* and *weakly meritocratic* fairness that holds *throughout the learning process*. This was inspired by Dwork et al. [5], who suggest fair treatment equates to treating "similar" people similarly,

*The full technical version of this paper is available at https://arxiv.org/abs/1610.09559.

where similarity is defined with respect to an assumed pre-specified task-specific metric. Taking the fairness formulation of Joseph et al. [9] as our starting point, our definition of fairness does not promise to correct for past inequities or inaccurate or biased data. Instead, it assumes the existence of an accurate mapping from features to true quality for the task at hand[1] and promises fairness while learning and using this mapping in the following sense: any *individual* who is currently more qualified (for a job, loan, or college acceptance) than another individual will always have at least as good a chance of selection as the less qualified individual.

The one-sided nature of this guarantee, as well as its formulation in terms of quality, leads to the name *weakly meritocratic* fairness. Weakly meritocratic fairness may then be interpreted as a minimal guarantee of fairness: an algorithm satisfying our fairness definition cannot favor a worse option but is not required to favor a better option. In this sense our fairness requirement encodes a necessary variant of fairness rather than a completely sufficient one.

We additionally note that our fairness guarantees require fairness *at every step of the learning process*. We view this as an important point, especially for algorithms whose learning processes may be long or continuous. Furthermore, while it may seem reasonable to relax this requirement to allow a small fraction of unfair steps, it is unclear how to do so without enabling discrimination against a correspondingly small population.

2 Model

Fix some $\beta \in [-1, 1]^d$, the underlying linear coefficients of our learning problem, and T the number of rounds. For each $t \in [T]$, let C_t be a convex body denoting the set of available choices in round t. An algorithm $\mathcal{A}$, facing choices C_t, picks a single $x_t \in C_t$, and observes reward y_t such that $\mathbb{E}[y_t] = \langle \beta, x_t \rangle$, and the distribution of the noise $\eta_t = y_t - \langle \beta, x_t \rangle$ is sub-Gaussian, i.e. has tails dominated by those of a Gaussian distribution. Let $\mathbf{X}_t = [X_1; \ldots; X_t], \mathbf{Y}_t = [Y_1; \ldots; Y_t]$ refer to the design and observation matrices at round t.

Regret The notion of regret we will consider is that of pseudo-regret. Facing a sequence of choice sets $C_1, \ldots, C_T$, suppose $\mathcal{A}$ chooses points $x_1, \ldots, x_T$.[2] Then, the expected reward of $\mathcal{A}$ on this sequence is $\mathrm{Rew}(\mathcal{A}) = \mathbb{E}\left[\sum_{t \in [T]} y_t\right]$.

[1] Friedler et al. [7] provide evidence that providing fairness from bias-corrupted data is quite difficult.

[2] If these are randomized choices, the randomness of $\mathcal{A}$ is incorporated into the expected value calculations.

Refer to the sequence of feasible choices which maximizes expected reward as $x_1^*, \ldots, x_T^*$, defined with full knowledge of β.

Then, the **pseudo-regret** of $\mathcal{A}$ on any sequence is defined as

$$\mathrm{Rew}(x_1^*, \ldots, x_T^*) - \mathrm{Rew}(\mathcal{A}) = R(T).$$

The **pseudo-regret** of $\mathcal{A}$ refers to the maximum pseudo-regret $\mathcal{A}$ incurs on any sequence of choice sets and any $\beta \in [-1, 1]^d$. If $R(T) = o(T)$, then $\mathcal{A}$ is said to be **no-regret**. If, for any input parameter $\delta > 0$, $R(T)$ upper-bounds the expectation of the rewards of the sequence chosen by $\mathcal{A}$ with probability $1 - \delta$, then we call this a *high-probability* regret bound for $\mathcal{A}$.

Fairness Consider an algorithm $\mathcal{A}$, which chooses a sequence of *probability distributions* $\pi_1, \pi_2, \ldots, \pi_T$ over feasible sets to pick, $\pi_t \in \Delta(2^{C_t})$. Note that distribution π_t depends upon $C_1, \ldots, C_t$, the choices $P_1, \ldots, P_{t-1}$, and $Y_1, \ldots, Y_{t-1}$.

We adapt our fairness definition from Joseph et al. [9], generalizing from discrete distributions over finite action sets to mixture distributions over possibly infinite action sets. Given an action space D, and an algorithm $\mathcal{A}$, let $\pi_t \in \Delta(D)$ be the distribution on actions by $\mathcal{A}$ at time t.

Definition 1 (Weakly Meritocratic Fairness)**.** We say that an algorithm $\mathcal{A}$ is *weakly meritocratic* if, for any input $\delta \in (0, 1]$ and for any θ, with probability at least $1 - \delta$, at every round t one of the following three conditions is satisfied, depending on the nature of π_t:

- If π_t is a discrete distribution: For $g_{ti}(x) = \pi_{ti}(x)$ (the probability mass function), for all x, y such that $\langle \theta, x \rangle \geq \langle \theta, y \rangle$,
$$g_{ti}(x) \geq g_{ti}(y).$$
- If π_t is a continuous distribution: For $g_{ti}(x) = f_{ti}(x)$ (the probability density function), for all x, y such that $\langle \theta, x \rangle \geq \langle \theta, y \rangle$,
$$g_{ti}(x) \geq g_{ti}(y).$$
- If π_t can be written as a mixture distribution: $\sum_i \alpha_i \pi_{ti}, \sum_i \alpha_i = 1$, such that each constituent distribution $\pi_{ti} \in \Delta(D)$ is either discrete or continuous and satisfies one of the above two conditions.

For brevity, since we do not consider other fairness notions in this paper, we will often refer to weakly meritocratic algorithms simply as "fair". We say $\mathcal{A}$ is **round-fair** at time t if π_t satisfies the above conditions.

This generalization allows for mixtures over distributions, some of which may be continuous and others discrete. This in particular allows an algorithm which is round-fair in round t to randomly choose between playing one of two fair distributions. For example, it would satisfy Definition 1 to play the uniquely best action in D with probability $\frac{1}{2}$, and with probability $\frac{1}{2}$ choose an action uniformly at random from D.

3 Fair algorithms for convex action sets

In this section we analyze linear bandits with infinite choice sets in the familiar 1-bandit setting.[3] We provide a fair algorithm with an instance-dependent sublinear regret bound for infinite choice sets – specifically convex bodies – below. In Section 4 we match this with lower bounds showing that instance dependence is an unavoidable cost for fair algorithms in an infinite setting.

A naive adaptation of RidgeFair to an infinite setting requires maintenance of infinitely many confidence intervals and is thus impractical. We instead assume that our choice sets are convex bodies and exploit the resulting geometry: since our underlying function is linear, it is maximized at an *extremal* point. This simplifies the problem, since we need only reason about the relative quality of extremal points. The relevant quantity is Δ_{gap}, a notion adapted from Dani et al. [4] that denotes the difference in reward between the best and second-best extremal points in the choice set. When Δ_{gap} is large it is easier to confidently identify the optimal choice and select it deterministically without violating fairness. When Δ_{gap} is small, it is more difficult to determine which of the top two points is best – and since deterministically selecting the wrong one violates fairness for any points infinitesimally close to the true best point, we must play randomly from the entire choice set.

Our resulting fair algorithm, FairGap, proceeds as follows: in each round it uses its current estimate of β to construct confidence intervals around the two choices with highest estimated reward and selects the higher one if these intervals do not overlap; otherwise, it selects uniformly at random from the entire convex body. We prove fairness and bound regret by analyzing the rate at which random exploration shrinks our confidence intervals and relating it to the frequency of exploitation, a function of Δ_{gap}. We begin by formally defining Δ_{gap} below.

[3]Note that no-regret guarantees are in general impossible for infinite choice sets in m-bandit and k-bandit settings, since the continuity of the infinite choice sets we consider makes selecting multiple choices while satisfying fairness impossible without choosing uniformly at random from the entire set.

Definition 2 (Gap, adapted from Dani et al. [4])**.** Given sequence of action sets $C = (C_1, \ldots, C_T)$, define Ω_t to be the set of extremal points of C_t, i.e. the points in C_t that cannot be expressed as a proper convex combination of other points in C_t, and let $x_t^* = \max_{x \in C_t} \langle \beta, x \rangle$. The *gap* of C_t is

$$\Delta_{\text{gap}} = \min_{1 \le t \le T} \left(\inf_{x_t \in \Omega_t, x_t \neq x_t^*} \langle \beta, x_t^* - x_t \rangle \right).$$

Δ_{gap} is a lower bound on difference in payoff between the optimal action and any other extremal action in any C_t. When $\Delta_{\text{gap}} > 0$, this implies the existence of a unique optimal action in each C_t. Our algorithm (implicitly) and our analysis (explicitly) exploits this quantity: a larger gap enables us to confidently identify the optimal action more quickly.

We now present the regret and fairness guarantees for FairGap.

Theorem 1. *Given sequence of action sets $C = (C_1, \ldots, C_T)$ where each C_t has nonzero Lebesgue measure and is contained in a ball of radius r and feedback with R-sub-Gaussian noise,* FairGap *is fair and achieves*

$$\textsc{Regret}(T) = O\left(\frac{r^6 R^2 \ln(2T/\delta)}{\kappa^2 \lambda^2 \Delta_{gap}^2} \right)$$

where $\kappa = 1 - r\sqrt{\frac{2\ln\left(\frac{2dT}{\delta}\right)}{T\lambda}}$ *and* $\lambda = \min_{1 \le t \le T} \left[\lambda_{\min}(\mathbb{E}_{x_t \sim_{UAR} C_t}[x_t^T x_t]) \right]$

We sketch the proof of Theorem 3 here: we first bound the influence of noise on the confidence intervals we construct (via matrix Chernoff bounds) and prove that, with high probability, FairGap constructs correct confidence intervals. This requires reasoning about the spectrum of the covariance matrix of each choice set, which is governed by λ, a quantity which, informally, measures how quickly we learn from uniformly random actions. [4]. With correct confidence intervals in hand, fairness follows almost immediately, and to bound regret we analyze the rate at which these confidence intervals shrink.

The analysis above implies identical regret and fairness guarantees when each C_t is finite. For comparison, the results of our companion paper guarantee $\textsc{Regret}(T) = O(dk\sqrt{T})$. This result, in comparison, enjoys a regret independent of k which may prove especially useful for cases involving large k.

Finally, our analysis so far has elided any computational efficiency issues arising from sampling randomly

[4]λ can be computed directly for finite C_t or approximated by any positive lower bound for infinite C_t and substituted directly into our results.

from C. We note that it is possible to circumvent this issue by relaxing our definition of fairness to *approximate fairness* and obtain similar regret bounds for an efficient implementation. We achieve this using results from the broad literature on sampling and estimating volume in convex bodies, as well as recent work on finding "2nd best" extremal solutions to linear programs.

4 Instance-dependent Lower Bound for Fair Algorithms

We now present a lower bound instance for which any fair algorithm *must* suffer gap-dependent regret. More formally, we show that when each choice set is a square, i.e. $C_t = [0,1]^2$ for all t, for any fair algorithm $\textsc{Regret}(T) = \tilde{\Omega}(1/\Delta_{\text{gap}})$ with probability at least $1-\delta$. This also implies the weaker result that no fair algorithm enjoys an instance-independent sub-linear regret bound $o(T)$ holding uniformly over all β. We therefore obtain a clear separation between fair learning and the unconstrained case [4], and show that an instance-dependent upper bound like the one in Section 3 is unavoidable. Our arguments establish fundamental constraints on fair learning with large choice sets and quantify through the Δ_{gap} parameter how choice set geometry can affect the performance of fair algorithms. The lower bound employs a Bayesian argument resembling that in [9] but with a novel "chaining" argument suited to infinite action sets. We present the result for $d = 2$ for simplicity; the proof technique holds in any dimension $d \geq 2$.

Theorem 2. *For all t let $C_t = [-1,1]^d$, $\beta \in [-1,1]^d$, and $y_t = \langle x_t, \beta\rangle + \eta_t$, where $\eta_t \sim U[-1,1]$. Let $\mathcal{A}$ be any fair algorithm. Then for every gap Δ_{gap}, there is a distribution over instances with gap $\Omega(\Delta_{gap})$ such that any fair algorithm has regret $\textsc{Regret}(T) = \tilde{\Omega}(1/\Delta_{gap})$ with probability $1-\delta$.*

We again sketch of the central ideas in the proof. We start with the fact that any fair algorithm $\mathcal{A}$ is required to be fair for any value β of the linear parameter. Thus if we draw $\beta \sim \tau$, $\mathcal{A}$ must be round-fair for all $t \geq 1$ with probability at least $1-\delta$, where the probability includes the random draw $\beta \sim \tau$. Then Bayes' rule implies that the procedure that draws $\beta \sim \tau$ and then plays according to $\mathcal{A}$ is identical to the procedure which at each step t re-draws β from its posterior distribution given the past $\tau|_{h_t}$.

Next, given the prior τ, $\mathcal{A}$'s round fairness at step t requires that (with high probability) if $\mathcal{A}$ plays action x with higher probability than action y, we must have

$$\mathbb{P}_{\beta\sim\tau|_{h_t}}\left[\langle\beta, x\rangle > \langle\beta, y\rangle\right] > \frac{3}{4}. \tag{1}$$

This enables us to reason about the fairness and regret of the algorithm via a specific analysis of the posterior distribution $\tau|_{h_t}$. This Bayesian trick, first applied in [9], is a general technique useful for proving fairness lower bounds.

We then show that for a choice of prior specific to our choice set C, that two things hold: (i) whenever $\tau|_{h_t} = \tau$, Equation 1 forces $\mathcal{A}$ to play uniformly from C, and (ii) with high probability $\tau = \tau|_{h_t}$ until $t > \tilde{\Omega}(1/\epsilon)$, where ϵ is a parameter of the prior that acts as a proxy for Δ_{gap}. Playing an action uniformly from C incurs $\Omega(1)$ regret per round, so these two facts combine to show that with high probability $\textsc{Regret}(T) = \tilde{\Omega}(1/\epsilon)$.

Finally we consider $\textsc{Regret}(T)$ conditional on the event that $\Delta_{\text{gap}}(\beta) > \delta \cdot \epsilon$, which by our construction of τ happens with probability $1-\delta$, and show that this with high probability implies $\textsc{Regret}(T) = \Omega\left(\frac{1}{\epsilon}\right)$. We then show that when $\beta \sim \tau_{gap}$, $\Delta_{\text{gap}}(\beta) \geq \delta\cdot\epsilon$. Thus, when $\beta \sim \tau_{gap}$, with high probability, $\textsc{Regret}(T) = \tilde{\Omega}(1/\epsilon) = \tilde{\Omega}(1/\Delta_{\text{gap}})$, as desired.

The proof uses the fact that when $\tau = \tau|_{h_t}$, Equation 1 forces $\mathcal{A}$ to play uniformly at random. This happens by transitivity: if Equation 1 forces $\mathcal{A}$ to play x equiprobably with y and y equiprobably with z, then x must be played equiprobably with z. Finally, we note that this impossibility result only holds for $d \geq 2$. When $d = 1$, the problem reduces to estimating the sign of β, which takes $O(1/\beta^2)$ observations of random play and accumulates $O(\sqrt{T})$ regret.

5 Zero Gap: Impossibility Result

Section 3 presents an algorithm for which the sublinear regret bound has dependence $1/\Delta_{\text{gap}}^2$ on the instance gap. Section 4 exhibits an choice set C with a $\tilde{\Omega}(1/\Delta_{\text{gap}})$ dependence on the gap parameter. We now exhibit a choice set C for which $\Delta_{\text{gap}} = 0$ for every β, and for which no fair algorithm can obtain non-trivial regret for any value of β. This precludes even instance-dependent fair regret bounds on this action space, in sharp contrast with the unconstrained bandit setting.

Theorem 3. *For all t let $C_t = S^1$, the unit circle, and $\eta_t \sim$ Unif$(-1,1)$. Then for any fair algorithm $\mathcal{A}$, $\forall\beta \in S^1, \forall T \geq 1$, we have*

$$\mathbb{E}_\beta[\textsc{Regret}(T)] = \Omega(T).$$

S^1 makes fair learning difficult for the following reasons: since S^1 has no extremal points, there is no finite set of points which for any β contains the uniquely optimal action, and for any point in S^1, and any finite set of observations, there is another point in S^1 for which the

algorithm cannot confidently determine relative reward. Since this property holds for *every* point, the fairness constraint transitively requires that the algorithm play every point uniformly at random, at every round.

References

[1] Richard Berk, Hoda Heidari, Shahin Jabbari, Michael Kearns, and Aaron Roth. Fairness in criminal justice risk assessments: The state of the art. *arXiv preprint arXiv:1703.09207*, 2017.

[2] Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv preprint arXiv:1703.00056*, 2017.

[3] Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. *arXiv preprint arXiv:1701.08230*, 2017.

[4] Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *COLT*, pages 355–366, 2008.

[5] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of ITCS 2012*, pages 214–226. ACM, 2012.

[6] Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

[7] Sorelle A. Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. On the (im)possibility of fairness. In *arXiv*, volume abs/1609.07236, 2016. URL `http://arxiv.org/abs/1609.07236`.

[8] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, volume abs/1610.02413, 2016. URL `http://arxiv.org/abs/1610.02413`.

[9] Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, pages 325–333, 2016.

[10] J. Kleinberg, S. Mullainathan, and M. Raghavan. Inherent trade-offs in the fair determination of risk scores. In *ITCS*, Jan 2017.

[11] Blake Woodworth, Suriya Gunasekar, Mesrob I Ohannessian, and Nathan Srebro. Learning non-discriminatory predictors. *arXiv preprint arXiv:1702.06081*, 2017.

[12] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi. Fairness beyond disparate treatment and disparate impact: Learning classification without disparate mistreatment. In *Proceedings of World Wide Web Conference*, 2017.