Planning to Fairly Allocate: Probabilistic Fairness in the Restless Bandit Setting

Christine Herlihy* Aviva Prins* Aravind Srinivasan John P. Dickerson

University of Maryland, College Park

Introduction

Figure 1. In the restless multi-armed bandit setting, select $k \ll N$ arms at each timestep t. Each arm evolves according to an action-dependent Markov Decision Process (MDP).

Find a probabilistic policy π^* that maximizes reward and enforces the budget and (new!) distributive fairness constraints.

$$\pi^* = \arg\max_{\pi \in \mathbb{R}^N} R^\pi(S) \text{ s.t. } \sum_i p_i = k \text{ and } \forall i,\ p_i \in [\ell, u]$$

The Whittle Index:

$$W(b_t^i) = \inf_m \left\{ m \mid V_m(b_t^i, a_t^i = 0) \geq V_m(b_t^i, a_t^i = 1) \right\}$$

$$V_m(b_t^i) = \max \begin{cases} m + r(b_t^i) + \beta V_m\left(b_{t+1}^i\right) & \textit{passive} \\ r(b_t^i) + \beta \left[b_t^i V_m\left(P_{1,1}^1\right) + (1 - b_t^i) V_m\left(P_{0,1}^1\right) \right] & \textit{active} \end{cases}$$

Why distributive fairness?

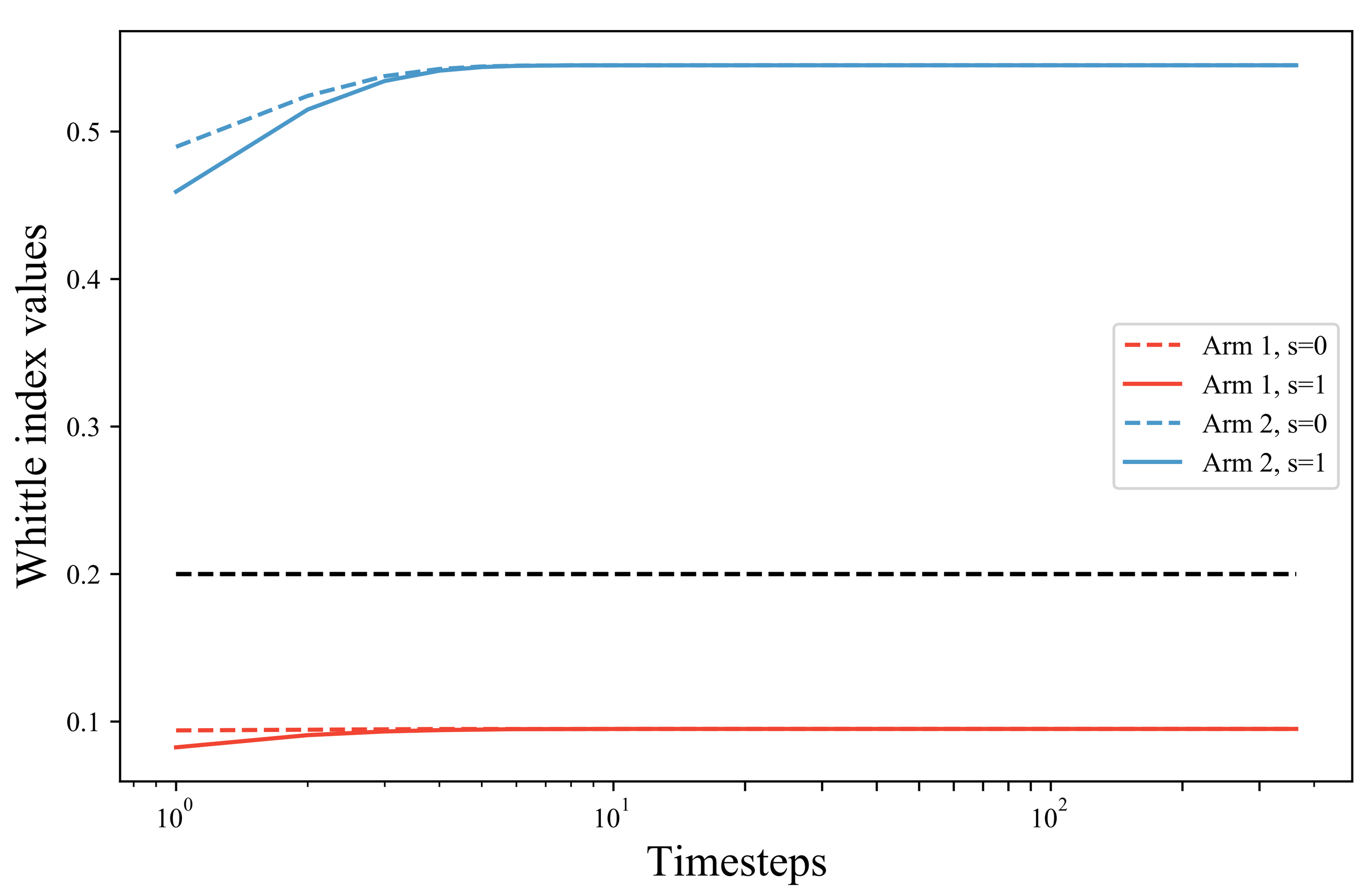
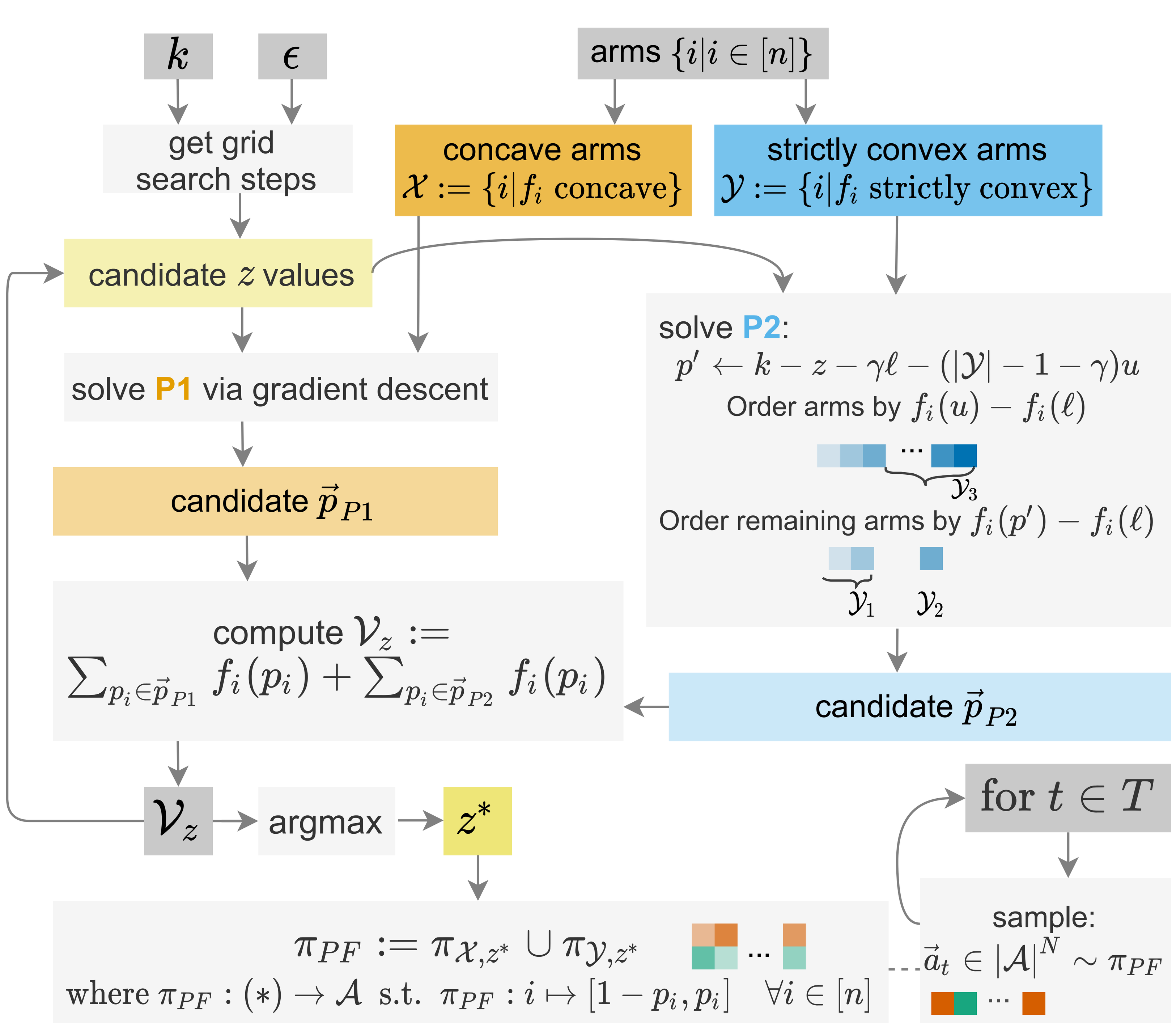


Figure 2. The Whittle index values for Arm 1 and 2 can be separated by a horizontal line, so (WLOG) Arm 2 will always be chosen over Arm 1 because its index value dominates.

ProbFair: a probabilistically fair policy



Experimental evaluation

Random§	Select k arms uniformly at random at each t.
Round-Robin§,‡	Select k arms at each t in fixed, sequential order.
TW-based heuristics‡	Select top-k arms based on Whittle index values. Available arms vary based on time-indexed fairness constraint satisfaction [3].
Risk-Aware TW (RA-TW)†	Select top-k arms based on Whittle index values, with a concave reward function [2].
Threshold Whittle (TW)*	Select top-k arms based on Whittle index values [4, 1].

Table 1. Comparison policies

$\min_i \mathbb{E}$[# pulls]	Policy		$\mathbb{E}$[IB] (%)	$\mathbb{E}$[EMD] (%)
10, $\ell = 0.056$, $\nu = 18$	PF	ℓ	88.45 $\pm$ 0.27	81.11 $\pm$ 0.18
	First	ν	88.75 $\pm$ 0.27	**68.19 $\pm$ 0.14**
	Last	ν	89.32 $\pm$ 0.26	69.17 $\pm$ 0.11
	Random	ν	**92.02 $\pm$ 0.18**	71.24 $\pm$ 0.13
18, $\ell = 0.1$, $\nu = 10$	PF	ℓ	81.57 $\pm$ 0.29	60.04 $\pm$ 0.22
	First	ν	81.07 $\pm$ 0.31	**47.44 $\pm$ 0.09**
	Last	ν	81.30 $\pm$ 0.29	48.47 $\pm$ 0.08
	Random	ν	**84.33 $\pm$ 0.26**	51.67 $\pm$ 0.10
30, $\ell = 0.167$, $\nu = 6$	PF	ℓ	68.22 $\pm$ 0.33	22.66 $\pm$ 0.17
	First	ν	70.22 $\pm$ 0.30	**19.10 $\pm$ 0.03**
	Last	ν	69.41 $\pm$ 0.33	19.70 $\pm$ 0.03
	Random	ν	**70.52 $\pm$ 0.34**	19.96 $\pm$ 0.04
comparison	TW		**100.00 $\pm$ 0.00**	**100.00 $\pm$ 0.00**
	RA-TW		72.73 $\pm$ 0.38	115.14 $\pm$ 0.26
baseline	Random		54.66 $\pm$ 0.35	10.44 $\pm$ 0.11
	NoAct		0.00 $\pm$ 0.00	76.08 $\pm$ 0.11
	RR		**62.96 $\pm$ 0.33**	**0.00 $\pm$ 0.00**

Table 2. $\mathbb{E}$[IB] and $\mathbb{E}$[EMD] by policy and fairness bracket

tl;dr: *Fairer* hyperparameters ($\ell \uparrow$, $\nu \downarrow$), yield decreased $\mathbb{E}$[IB] and $\mathbb{E}$[EMD], reflecting improved individual fairness at the expense of total reward. For each (ℓ, ν), ProbFair performs competitively with respect to the best-performing heuristic (which, like TW, are state-aware).

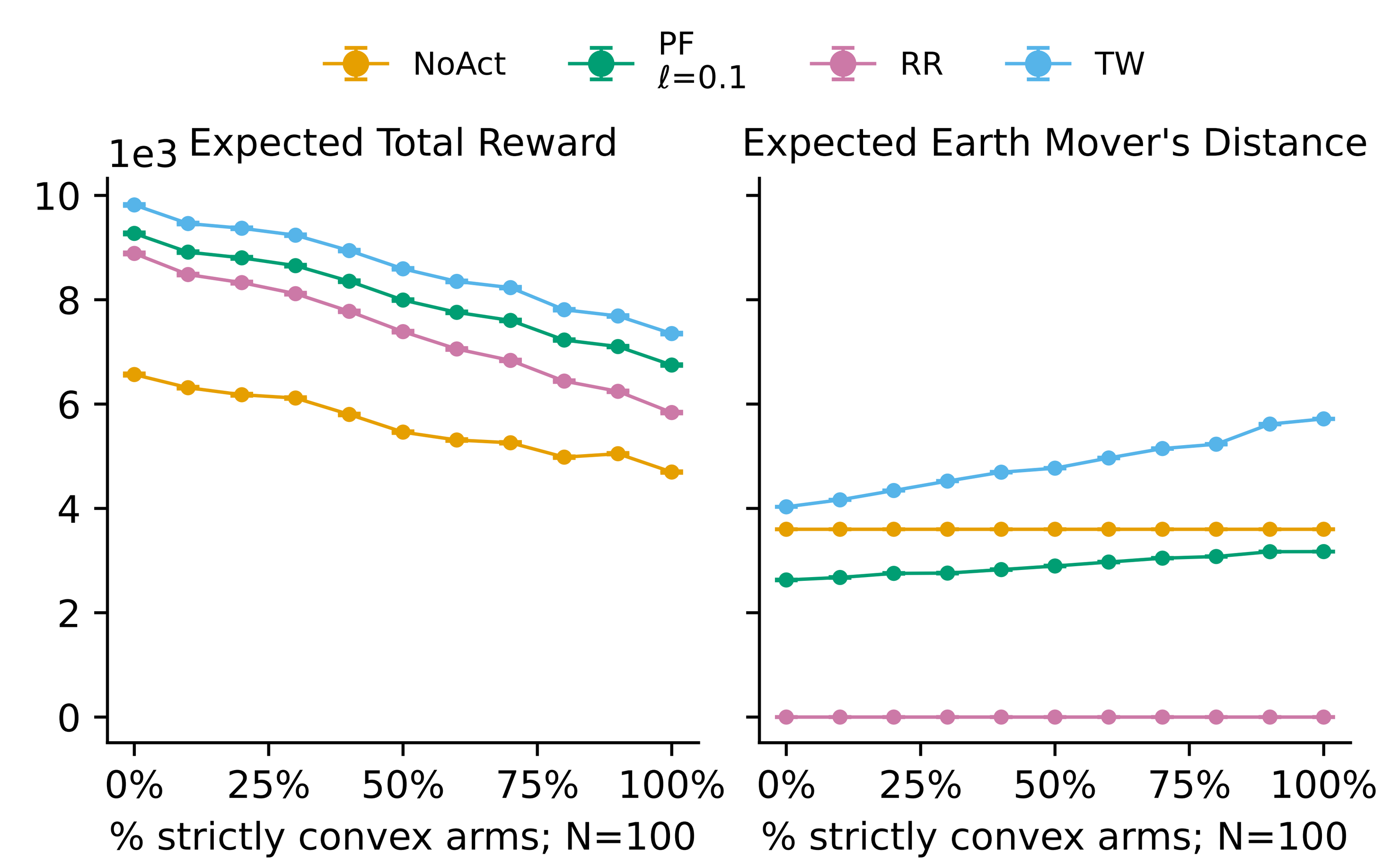


Figure 4. ProbFair evaluated on a breadth of randomly-generated cohorts.

tl;dr: $\mathbb{E}$[R] predictably declines for all policies as the % of unfavorable arms increases, while $\mathbb{E}$[EMD] rises for TW and ProbFair. ProbFair's *normalized* performance remains stable even as cohort composition is varied.

References

[1] A. Mate, J. Killian, H. Xu, A. Perrault, and M. Tambe. Collapsing Bandits and Their Application to Public Health Intervention. *Advances in Neural Information Processing Systems (NeurIPS)*, 33, 2020.

[2] A. Mate, A. Perrault, and M. Tambe. Risk-Aware Interventions in Public Health: Planning with Restless Multi-Armed Bandits. In *20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, London, UK, 2021.

[3] A. Prins, A. Mate, J. A. Killian, R. Abebe, and M. Tambe. Incorporating Healthcare Motivated Constraints in Restless Bandit Based Resource Allocation. *preprint*, 2020.

[4] P. Whittle. Restless Bandits: Activity Allocation in a Changing World. *Journal of Applied Probability*, 25(A):287–298, 1988.

Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO) 2022

avivaprins.com/probfair