Uniformly Bounded Regret in Dynamic Fair Allocation

Santiago R. Balseiro, Shangzhou Xia

Graduate School of Business, Columbia University, New York, NY 10027, srb2155@columbia.edu, sx2182@columbia.edu

We study a dynamic allocation problem in which $T$ sequentially arriving divisible resources are to be allocated to a number of agents with linear utilities. The marginal utilities of each resource to the agents are drawn stochastically from a known joint distribution, independently and identically across time, and the central planner makes immediate and irrevocable allocation decisions. Most works on dynamic resource allocation aim to maximize the utilitarian welfare, i.e., the efficiency of the allocation, which may result in unfair concentration of resources on certain high-utility agents while leaving others’ demands under-fulfilled. In this paper, aiming at balancing efficiency and fairness, we instead consider a broad collection of welfare metrics, the Hölder means, which includes the Nash social welfare and the egalitarian welfare. To this end, we first study a fluid-based policy derived from a deterministic surrogate to the underlying problem and show that for all smooth Hölder mean welfare metrics it attains an $O(1)$ regret over the time horizon length $T$ against the hindsight optimum, i.e., the optimal welfare if all utilities were known in advance of deciding on allocations. However, when evaluated under the non-smooth egalitarian welfare, the fluid-based policy attains a regret of order $\Theta(\sqrt{T})$. We then propose a new policy built thereupon, called Backward Infrequent Re-solving with Thresholding (BIRT), which consists of re-solving the deterministic surrogate problem at most $O(\log \log T)$ times. We prove the BIRT policy attains an $O(1)$ regret against the hindsight optimal egalitarian welfare, independently of the time horizon length $T$. We conclude by presenting numerical experiments to corroborate our theoretical claims and to illustrate the significant performance improvement against several benchmark policies.

1. Introduction

Fair resource allocation has been widely studied in economics and operations research, with applications in school seat allocation (Abdulkadiroğlu and Sönmez 2003, Abdulkadiroğlu et al. 2017), work visa lotteries (Pathak et al. 2020), affordable housing allocation (Kaplan 1984,
Arnosti and Shi (2020), machine load balancing (Azar et al. 1995, Azar 1998), etc. Involving multiple parties stipulates fairness as a crucial concern to the central planner. To this end, allocations are often evaluated on a balance between efficiency and fairness across the agents (Bertsimas et al. 2011, 2012, Ahmed et al. 2017, Donahue and Kleinberg 2020, Zeng and Psomas 2020). Common fairness criteria studied in the literature of fair allocation include Pareto efficiency (Varian 1974), envy-freeness (Abdulkadiroglu et al. 2017) along with its variants, and proportional fairness (Brandt et al. 2016, Banerjee et al. 2022b).

Cardinal welfare metrics incorporate efficiency and fairness concerns of allocations by mapping agents’ individual utilities to a real-valued numeric representation of collective welfare. These metrics are wieldy in operations research and computer science because they allow to cast allocation problems as optimization problems, and different allocations may be compared against one another, yielding a quantifiable welfare difference. One particular cardinal fairness metric is the egalitarian welfare, based on the theory of distributive justice by Rawls (2001, 2020), whose difference principle maximizes the welfare of the worst-off group of agents. The egalitarian welfare objective defines a max-min welfare optimization problem, whose indivisible variant is dubbed the Santa Claus problem (Bansal and Sviridenko 2006, BEN-ARI 1998, Chakrabarty et al. 2009). Another well-studied cardinal example is the Nash social welfare, mathematically the geometric mean of the agents’ utilities (Nash Jr. 1950, Luce and Raiffa 1989, Hahn et al. 1982, Freeman et al. 2020, Caragiannis et al. 2019, Conitzer et al. 2017). Optimal allocations under the Nash social welfare are invariant to scaling of any single agent, for it is homogeneous on each agent’s utility. In this work we consider a broad parameterized collection of Hölder mean welfare metrics (Wu et al. 2021), which subsumes the two examples above.

Allocation problems in real life, however, often concern dynamic and irrevocable decisions where either agents or resources arrive in a sequential manner (Nanda et al. 2020, Dickerson et al. 2019, Cheung et al. 2020). Some literature has referred to this variant as the online rationing problem, with applications in allocating computational resources to cloud users (Dinh et al. 2020,
This work focuses on the specific variant of allocating resources arriving sequentially to heterogeneous agents, which is particularly relevant when resources are perishable or urgently needed and require immediate allocation decisions. Applications include allocating donated organs to hospitals (Agarwal et al. 2019), donated food to local charities (Aleksandrov et al. 2015, Lien et al. 2014), and online advertisement slots to advertisers (Choi et al. 2020, Watts 2021, Gollapudi and Panigrahi 2014, Li et al. 2021, Bateni et al. 2022, Arnosti et al. 2016), etc. Our objective is to design computationally efficient dynamic policies that maximize the overall welfare.

Contributions In this work, we consider the dynamic fair allocation problem, where $T$ resources arriving sequentially are to be immediately and irrevocably allocated to $n$ agents. The joint marginal utilities of each resource to the agents are drawn independently from a joint distribution that is known to the central planner, which are revealed before allocation decisions are made. An optimal online policy could in principle be computed using dynamic programming but, because of the so-called curse of dimensionality, solving the problem to optimality is impractical when the time horizon is large. We therefore seek to design policies that are computationally efficient and have provable performance guarantees.

We first consider a static policy that is obtained by solving a fluid relaxation in which all stochastic quantities are replaced by their mean values. Simple as it is, we show under all smooth Hölder-mean welfare metrics it attains a resounding $O(1)$ regret over any $T$ against the hindsight optimal solution, i.e., the optimal allocation if all marginal utilities were known in advance. The sole non-smooth exception is the egalitarian welfare objective, under which fluid policies attain a regret on the order of $\Theta(\sqrt{T})$. Refer to Section 3 for rigorous definitions of the fluid relaxation problem, its optimum $\text{FLU}$, and fluid policies $F$.

Our next main contribution is proposing the Backward Infrequent Re-solving with Thresholding (BIRT) policy for dynamic fair allocation under the egalitarian welfare. Our policy is inspired
by Bumpensanti and Wang (2020), who designed similar policies for utilitarian objectives based on fluid policies, and consists of updating and re-solving the fluid problem at most $O(\log \log T)$ times. Utilitarian objectives are simpler to handle because they are separable over time, i.e., the total utility is simply the sum of the utilities of each time period. By contrast, the egalitarian objective discussed in our paper is non-linear and not time separable, and so requires a different analysis. We provide a novel analysis to show that the BIRT policy achieves $O(1)$ regret against the hindsight optimum, namely uniformly bounded over the time horizon length $T$. The uniformly bounded regret guarantee implies that the BIRT policy performs almost comparably to a clairvoyant who foresees all arrivals in advance of acting.

A salient feature of the BIRT policy is that its $O(1)$ regret guarantee is insensitive to whether or not the original fluid problem is degenerate (more technically, that any fluid policy is non-degenerate) or nearly degenerate. This is in drastic contrast to previous works on online stochastic optimization whose similar $O(1)$-loss results rely heavily on non-degeneracy conditions on the underlying fluid problem (see Section 5 for a numerical illustration).

$$q = -\infty \quad \text{egalitarian} \quad \text{Expected welfare} \quad q \in (-\infty, 1] \quad \text{Hölder-mean}$$

| Lemma 2 | EC.9 | $\Theta(\sqrt{T})$ |
|---|---|---|
| FLU |
| Lemma 1, 4 | $\Theta(\sqrt{T})$ |
| OPT |
| BIRT |
| F |
| $O(1)$ | Theorem 1 |

**Figure 1** Summary of benchmark values and policy performances. Results relevant to the egalitarian welfare are shown on the left, and that relevant to all other Hölder-mean welfare metrics is shown on the right.

We summarize findings on the benchmarks (FLU and OPT) and policy performances (BIRT and F) on the vertical axis in Figure 1 with respective suprema of the differences (i.e., worst-case arrival distribution) between them. All in all, we provide computationally efficient online policies that attain uniformly bounded regret in for all Hölder-mean welfare metrics.
1.1. Related work

Our work contributes to the literature on fair resource allocation and, more generally, to the literature on dynamic resource allocation problems.

*Fair allocation* There is a stream of literature studying the online version of the problem in which the arrival process in unknown to the central planner. Devanur et al. (2019) consider the egalitarian objective when the agents’ utilities are drawn independently from a distribution that is unknown to the central planner (the “unknown IID” model) and design an algorithm that attains $O(\sqrt{T})$ regret. Agrawal and Devanur (2014) consider general online allocation problems with concave non-separable objectives, which subsume our problem with Hölder-mean welfare metrics, and propose an algorithm based on multiplicative weight updates method (Arora et al. 2012) that also attains $O(\sqrt{T})$ regret. Balseiro et al. (2021b) study similar general online allocation with hard resource constraints and propose fast algorithms, which do not require solving auxiliary optimization problems, based on mirror descent. More recently, Kawase and Sumita (2021) consider the max-min fair allocation of indivisible resources arriving online to heterogeneous agents. They propose a deterministic policy that achieves an asymptotic approximation ratio of $n$ in the adversarial arrival model, which they show is optimal. In the case of unknown IID arrivals, they give a policy that achieves $O(\sqrt{T})$ regret against the hindsight optimum benchmark, based on the multiplicative weights updates (Arora et al. 2012).

In contrast to the previous line of work, we consider the case where the distribution of utilities is known to the central planner. The $O(1)$ regret bound of the BIRT policy in the known IID model significantly improves upon the unknown IID model, in which the best known attainable regret is $O(\sqrt{T})$ (Agrawal and Devanur 2014, Devanur et al. 2019, Kawase and Sumita 2021, Balseiro et al. 2021b). The drastic performance improvement shows the great use of distributional knowledge. This is similar in spirit to Banerjee et al. (2022a), who show in maximizing Nash social welfare, agents’ utility predictions significantly improve the approximation ratio of online algorithms in adversarial settings to $O(\log(n \wedge T))$ from a trivial $\Theta(n)$. Other works on maximizing Nash social welfare include Freeman et al. (2017), Sinclair et al. (2020).
Under the egalitarian welfare, the particular case of identical agents has been studied for decades as the “online machine covering” problem in the context of scheduling (Deuermeyer et al. 1982, Azar and Epstein 1998, Gálvez et al. 2020, Woeginger 1997). Moreover, there is a line of work studying an offline version of the max-min fair allocation problem (the case where all utilities are known in advance), which is referred to as the Santa Claus problem (Li et al. 2019, Asadpour and Saberi 2010, Kawase and Sumita 2020, Bezáková and Dani 2005, Golovin 2005, Feige 2008, Chakrabarty et al. 2009, Haeupler et al. 2011).

In this paper, we study an online supply version of the dynamic fair allocation problem in which resources arrive online and the population of agents are fixed. The online demand version of the problem in which resources are fixed and agents arrive online has also been studied in the literature (Nanda et al. 2020, Manshadi et al. 2021, Ma et al. 2020, Lien et al. 2014).

Finally, there is a long line of literature on online allocation focusing on other fairness notions such as envy (Zeng and Psomas 2020, Benade et al. 2018), envy-freeness (up to 1 item) (He et al. 2019, Aleksandrov et al. 2015), Pareto efficiency (Zeng and Psomas 2020) and more general concave returns (Devanur and Jain 2012). We recommend to the reader the recent survey by Aleksandrov and Walsh (2020) for an overview of the literature of variants of online fair allocation problems.

Dynamic resource allocation Our work is also closely related methodologically to the stream of literature on dynamic resource allocation, especially revenue management, that aims at maximizing revenue or efficiency of an allocation given limited resources. Our policy is based on re-solving the fluid problem, a specific form of certainty equivalent of dynamic stochastic optimization problems when the problem instances are independent and identically distributed according to a distribution known to the central planner. The main difference with this line of work is that the Hölder-mean welfare objective we consider is not time separable, which calls for a different theoretical analysis.

The seminal works by Gallego and Van Ryzin (1997) and Talluri and Van Ryzin (1998) show fluid policies based on solving the fluid problem only once at the beginning suffice to achieve an
\(O(\sqrt{T})\) regret against the hindsight optimum. Reiman and Wang (2008) propose a policy based on re-solving the fluid policy exactly once only incurring an \(o(\sqrt{T})\) asymptotic revenue loss. Resolving multiple times or even every period can give an \(O(1)\) loss under the assumption that the fluid problem admits a nondegenerate optimal solution (Jasin and Kumar 2012, Wu et al. 2015). Bumpensanti and Wang (2020) propose the Infrequent Re-solving with Thresholding policy, which incurs an \(O(1)\) loss, and dispenses with the nondegeneracy condition. Our policy is based thereupon. Most recently, Vera and Banerjee (2021) and Vera et al. (2021) propose a general Bayes selector policy that attains low regret for a number of dynamic allocation problems. A specific variant called the multi-secretary problem also witnessed policies incurring \(O(1)\) (Arlotto and Gurvich 2019) or \(O(\log T)\) (Bray 2019) regret. We recommend to the reader a recent survey by Balseiro et al. (2021a) for an overview of dynamic allocation problems.

1.2. Notation

We denote the extended real line by \(\overline{\mathbb{R}} := \mathbb{R} \cup \{-\infty, +\infty\}\). For \(m \in \mathbb{N}\), we denote \([m] := \{1, 2, \ldots, m\}\), the probability \(m\)-simplex by \(\Delta_m := \{x \in \mathbb{R}_+^m : \|x\|_1 = 1\}\), a column \(m\)-vector of ones by \(1_m\) and a column \(m\)-vector of zeros by \(0_m\). For \(w \in \mathbb{R}\), we denote by \([w] := \max\{m \in \mathbb{Z} : m \leq w\}\) the largest integer smaller than or equal to \(w\), and by \([w] := \min\{m \in \mathbb{Z} : m \geq w\}\) the smallest integer larger than or equal to \(w\). For \(u, v \in \mathbb{R}^n\), we denote the element-wise product by \(u \ast v := (u^i v^i)\).

We use boldface alphabets (e.g., \(\xi, x\)) to denote matrices. For any two positive-valued functions \(f, g : \mathbb{N} \rightarrow \mathbb{R}_+\), we denote \(f(T) = o(g(T))\) if \(\lim_{T \rightarrow \infty} f(T)/g(T) = 0\); we denote \(f(T) = O(g(T))\) if \(\limsup_{T \rightarrow \infty} f(T)/g(T) < \infty\); we denote \(f(T) = \Omega(g(T))\) if \(\liminf_{T \rightarrow \infty} f(T)/g(T) > 0\); we denote \(f(T) = \Theta(g(T))\) if \(f(T) = \Omega(g(T))\) and \(f(T) = O(g(T))\). With a slight abuse of notation, we sometimes denote optimization problems and their optima interchangeably.

2. Problem formulation

In this work, we consider the problem of dynamically allocating resources to \(n\) heterogeneous agents with linear utilities that are stochastically distributed. More formally, suppose each agent
$i \in [n]$ starts with zero initial utility $B^i_0 = 0$ and passively accumulates linear utility while receiving allocated resources over time. The central planner receives a sequence of $T$ divisible resources, each along with a revealed vector $b_t \in \mathbb{R}^n_+$ indicating its joint marginal utilities to the $n$ agents, and immediately decides on its proportional allocation $x_t \in \Delta_n$ to the agents. In the sequel, we refer to the $t$-th resource and the $t$-th period interchangeably, and to resources and arrivals interchangeably. By the end of the time horizon, agent $i$ would have a cumulative utility of $B^i_T = \sum_{t=1}^T b^i_t x^i_t$.

We evaluate allocation decisions on their balance between efficiency and fairness through a cardinal welfare metric $w : \mathbb{R}^n_+ \to \mathbb{R}_+$ of the vector of agents’ cumulative utilities $B_T \in \mathbb{R}^n_+$. We denote the welfare generated by a sequence of allocation decisions $\mathbf{x} \equiv (x_t : t \in [T])$ according to policy $\pi$ by

$$\text{ALG}(\pi) := w \left( \sum_{t=1}^T b^i_t x^i_t \right). \quad (1)$$

An online policy is non-anticipative, or in other words, decides on allocations only based on the history observed so far. More formally, an online policy allocates $x_t \in \sigma(b_1, \ldots, b_t)$ for all $t \in [T]$, where $\sigma(b_1, \ldots, b_t)$ is the $\sigma$-algebra generated by the sequence of utility vectors. Notice we assume that online policies are deterministic for the purpose of evaluating the welfare they generated—this is without loss of optimality because deterministic online policies weakly dominate their randomized counterparts (Lemma EC.6).

A conventional benchmark used to evaluate online policies is the hindsight optimum, defined as the maximum welfare (1) if allocations $\mathbf{x}$ were allowed to be anticipative, i.e., all vectors of joint marginal utilities $(b_t : t \in [T])$ were known in advance. We denote by $\text{OPT}$ the hindsight optimum, as well as any such optimal policy with a slight abuse of notation. Formally, the hindsight optimum is the optimal value of the maximization problem

$$\text{OPT} := \max \left\{ w \left( \sum_{t=1}^T b^i_t x^i_t \right) : \mathbf{x} \in \Delta^T_n \right\}. \quad (2)$$

### 2.1. Welfare metric

In this work, we evaluate allocations using cardinal welfare metrics $w$, i.e., real-valued functions of the agents’ cumulative utilities $B_T \in \mathbb{R}^n_+$, which allow to balance between efficiency and fairness.
concerns. Hence, allocation decisions can be directly compared against one another, and their difference indicates a quantifiable welfare loss.

We next enumerate three cardinal welfare examples and briefly discuss the intuition behind them. The \textit{utilitarian welfare} \( \sum_{i \in [n]} B_i T \) neglects any fairness concerns; optimizing it leads to efficient allocations that could and most often lead to highly unfair concentration of resources at certain agents or groups. The \textit{egalitarian welfare} \( \min_{i \in [n]} B_i T \) is the other extreme—defined as the utility of the worst-off agent, this metric provides all agents with a uniform minimal utility guarantee. The \textit{Nash social welfare} \( (\prod_{i=1}^{n} B_i^q)^{1/n} \) is a more balanced metric in terms of the trade-off between efficiency and fairness; in addition, its geometric mean form satisfies positive homogeneity in \textit{each agent’s utility}, which insulates the central planner against agents’ possible strategic reporting of marginal utilities.

Formally, we consider a collection of welfare metrics that subsumes and generalizes the examples above, the \textit{Hölder means} \( w_q : \mathbb{R}_+^n \rightarrow \mathbb{R}_+ \), a.k.a. \textit{generalized power means}, parameterized by \( q \in \mathbb{R} \), defined as follows and illustrated in the fairness spectrum shown in Figure 2. In this problem, we stipulate \( q \in [-\infty, 1] \) because \( w_q \) for \( q \in (1, +\infty] \) is not concave and hence deemed “unfair” as per Lemma 1.

- For \( q = -\infty \), \( w_q(B) = \min_i B^i \);
- for \( q \in (-\infty, 0) \), \( w_q(B) = \left( \frac{1}{n} \sum_{i \in [n]} (B^i)^q \right)^{1/q} \) if \( B > 0 \) and \( w_q(B) = 0 \) otherwise;
- for \( q = 0 \), \( w_q(B) = \left( \prod_{i=1}^{n} B_i \right)^{1/n} \);
- for \( q \in (0, +\infty) \), \( w_q(B) = \left( \frac{1}{n} \sum_{i \in [n]} (B^i)^q \right)^{1/q} \);
- for \( q = +\infty \), \( w_q(B) = \max_i B^i \).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fairness_spectrum.png}
\caption{Fairness spectrum of the Hölder-mean welfare metrics \( w_q \) parameterized by \( q \in \mathbb{R} \). The fairness of the welfare metric decreases as \( q \) increases. Dashed line indicates the “unfair” regime \( (q > 1) \).}
\end{figure}
We consider the broad collection of Hölder mean welfare metrics \( w_q \) because they satisfy a number of axioms that are reasonable in common scenarios of fair allocation; moreover, classical results have shown that they are the only welfare metrics satisfying these axioms, up to a multiplicative constant, which we have disposed of through normalization of \( w_q \) in the definition above. Hence, \( w_q(u1) = u \) for any scalar \( u \geq 0 \), i.e., the welfare equals each agent’s utility if all utilities are identical.

**Lemma 1** ([Moulin (2004), Hardy et al. (1952)]) For any welfare metric \( w : \mathbb{R}_+^n \rightarrow \mathbb{R} \), \( w \propto w_q \) for some \( q \in [-\infty, 1] \) if and only if \( w \) satisfies the following axioms.

1. **Monotonicity:** \( w(B) \geq w(\tilde{B}) \) if \( B \geq \tilde{B} \).
2. **Symmetry:** \( w(B) = w(\sigma(B)) \) for any permutation \( \sigma \in \mathcal{S}_n \).
3. **Continuity:** \( w \) is continuous.
4. **Independence of unconcerned agents:** if \( w(B^1, B^2, \ldots, B^n) \geq w(\tilde{B}^1, B^2, \ldots, B^n) \), then \( w(B^1, \tilde{B}^2, \ldots, \tilde{B}^n) \geq w(\tilde{B}^1, B^2, \ldots, \tilde{B}^n) \).
5. **Homogeneity:** \( w(\lambda B) = \lambda w(B) \) for \( B \geq 0 \) and \( \lambda > 0 \).
6. **Pigou–Dalton principle:** \( w(B^1, B^2, B^3, \ldots) \geq w(B^1 + \varepsilon/2, B^2 - \varepsilon/2, B^3, \ldots) \) for \( \varepsilon \in (0, B^2 - B^1) \) if \( B^1 < B^2 \).

Note the axioms serve as the foundation to fair allocation and are reasonably motivated. More specifically, monotonicity establishes that larger agents’ utilities are preferable; symmetry guarantees equal treatment by making the welfare metric independent of the identities of the agents; independence of unconcerned agents determines that an agent’s preferences do not change if we modify the utility of other agents; homogeneity guarantees that allocation is independent of the units used to measured the utility; the Pigou-Dalton principle stipulates that the welfare metric prefers fairness to unfairness, i.e., transferring utility from a high-utility to a low-utility individual should increase fairness. By imposing these axioms, we restrict our attention to a class of welfare metrics that are rich enough to incorporate flexibility in fairness concerns and structured enough to work with.
2.2. Arrival model

We consider the IID arrival model. More formally, we assume marginal utility vectors $b_t$ are distributed according to some known finite joint distribution $P$ in the class $\mathcal{P}$, independently across time $t$. While our model assume that the joint marginal utilities are IID across time, they may be arbitrarily correlated across agents in a given period, allowing for heterogeneity in agents to the fullest extent.

Definition 1. The class of admissible probability distributions $\mathcal{P}$ is the set of all finite probability distributions $P$ with support in $[0, 1]^n$.

To be more precise, a distribution $P$ is admissible if it has a finite support $\{\beta_\ell \in [0, 1]^n : \ell \in [L]\}$ and is given by $P\{b = \beta_\ell\} = p_\ell$, where $L \in \mathbb{N}$ is the number of distinct arrival types, and $(p_\ell)_{\ell \in [L]} \in \Delta_L$ is a probability vector. In the sequel, the underlying distribution $P$ is omitted when implied in the context. We denote the number of type-$\ell$ arrivals by $N_\ell := \sum_{t=1}^{T} 1\{b_t = \beta_\ell\}$, which then follows a multinomial distribution, i.e., $(N_\ell : \ell \in [L]) \sim \text{Multinomial}(T; (p_\ell : \ell \in [L]))$ with $T$ trials and success probabilities $(p_\ell : \ell \in [L])$.

We evaluate an online policy using its expected H"{o}lder-mean welfare loss against the hindsight optimum, along with the relative counterpart. More rigorously, given underlying arrival distribution $P \in \mathcal{P}$, we define the regret and relative regret of a policy $\pi$ by

$$R_T(\pi) := \mathbb{E}[\text{OPT} - \text{ALG}(\pi)] \quad \text{and} \quad \rho_T(\pi) := \frac{\mathbb{E}[\text{OPT} - \text{ALG}(\pi)]}{\mathbb{E}[\text{OPT}]}$$

if $\mathbb{E}[\text{OPT}] > 0$ and $\rho_T(\pi) := 0$ otherwise.

3. Fluid problem and static policies

Certainty equivalents have been widely used in general dynamic stochastic optimization problems to design simple heuristics with good performance guarantees (Bertsekas 2012). Certainty equivalent policies seek to approximate the offline optimization by replacing stochastic quantities with fixed and known deterministic values. A particular choice is to replace all random quantities by their means. The resulting optimization problem is usually referred to as the fluid problem, or the
deterministic program. The optimal solution to the fluid problem prescribes a simple static policy, which can be implemented in the original stochastic program.

Before moving forward, we formally define the collection of static policies, which will play an important role in our analysis. A static policy specifies a proportional allocation vector \( \xi_\ell \in \Delta_n \) over agents for each type \( \ell \in [L] \) and allocates accordingly \( x_t = \sum_{\ell \in [L]} \xi_\ell 1\{b_t = \beta_\ell\} \). With an abuse of notation, we also refer to the underlying matrix \( \xi \equiv (\xi_\ell : \ell \in [L]) \) as a static policy. In other words, a static policy is one that allocates incoming resources only depending on their types, i.e., it makes identical allocation decisions for resources of the same type. Under static policy \( \xi \), the cumulative utility of agent \( i \) is \( B^i_T = \sum_{\ell \in [L]} N_\ell \beta^i_\ell \xi^i_\ell \). It is worth noting that any hindsight optimal policy can always be implemented as a static policy (see Lemma EC.7 in Appendix EC.2).

We now formally define the fluid problem. Consider the IID arrival model with distribution \( P \in \mathcal{P} \) known to the online central planner. In the fluid problem, the central planner assumes that the number of type-\( \ell \) arrivals is exactly \( \mathbb{E}[N_\ell] = T p_\ell \). Because there is no uncertainty, it is easy to see that static policies are optimal, and the fluid problem is defined as

\[
\text{FLU} := \max \left\{ w \left( T \sum_{\ell \in [L]} p_\ell \beta^i_\ell \xi^i_\ell \right) : \xi \in \Delta_n^L \right\}. \tag{4}
\]

In the fluid problem, the central planner chooses an static policy that maximizes the welfare corresponding to the agents’ expected utilities, where that of agent \( i \) under the static policy \( \xi \) is \( \mathbb{E}[B^i_T] = T \sum_{\ell \in [L]} p_\ell \beta^i_\ell \xi^i_\ell \). With an abuse of notation, we also denote by FLU its optimum value, a.k.a. the fluid benchmark.

An optimal solution to the fluid problem (4) is called a fluid (static) policy, denoted by \( \xi^F \) or \( F \). Fluid policies are classical heuristic policies that serve as a powerful tool for the online central planner who has no other knowledge than the underlying arrival distribution. The pseudocode for fluid policies is given in Algorithm [1].

The positive homogeneity of the welfare metric \( w \) implies that fluid policies are in fact independent of the time horizon length \( T \) and that the fluid benchmark FLU is simply a linear function of \( T \).
Algorithm 1: Fluid Policy (F)

**Input:** time horizon length $T \in \mathbb{N}$, arrival distribution $P \in \mathcal{P}$.

**Initialize:** initial utilities $B_i^0 \leftarrow 0$ for $i \in [n]$;

solve the fluid problem with initial utilities:

$$\xi^F \in \arg\max \left\{ w \left( T \sum_{\ell \in [L]} p_\ell \beta_\ell \xi^F_\ell \right) : \xi_\ell \in \Delta_n^L \right\}$$

for $t = 1, \ldots, T$ do // implement the fluid policy

observe incoming supply with utility vector $b_t$;

if $b_t = \beta_\ell$ then

act $x_i \leftarrow (\xi^F)_\ell$;

update $B_i^t \leftarrow B_i^{t-1} + b_i^t x_i^t$ for every agent $i \in [n]$;

**Output:** allocations $(x_i : t = 1, 2, \ldots, T)$ and welfare $w(B_T)$.

The classical literature in revenue management have shown that for utilitarian welfare objectives, fluid policies incur a regret at most on the order of $\sqrt{T}$ (Gallego and Van Ryzin 1997, Talluri and Van Ryzin 1998), which is tight in the worst case. Here we present similar results for the fluid benchmark and fluid policies in our dynamic fair allocation problem.

**Proposition 1.** Fix $P \in \mathcal{P}$. Under any Hölder-mean welfare $w_q$ with $q \in [-\infty, 1]$, the regret of any fluid policy $F$ is bounded by $R_T(F) = O(\sqrt{T})$.

Proposition 1 is in fact a corollary of the two following lemmata. Lemma 2 establishes the fluid benchmark as an upper bound of the expected hindsight optimum. The result is standard and follows from Jensen’s inequality and the concavity of the welfare metric (see proof in Appendix EC.1.1). Hence, the regret of any policy (e.g., a fluid policy) can be bounded by its expected welfare loss from the fluid benchmark.

**Lemma 2.** For any $P \in \mathcal{P}$, under any Hölder-mean welfare $w_q$ with $q \in [-\infty, 1]$, $\mathbb{E}[\text{OPT}] \leq \text{FLU}$.

The next result, Lemma 3, establishes a stochasticity gap at most of order $\sqrt{T}$ between the welfare corresponding to the agents’ expected utilities and the expected welfare under any static policy,
following from a concentration of measure argument. The proof can be found in Appendix EC.1.2. Proposition 1 follows because $\text{FLU} = w(E[B_T])$ and $\text{ALG}(F) = w(B_T)$ under the fluid policy $F$.

**Lemma 3.** Fix $P \in \mathcal{P}$. Under any Hölder-mean welfare $w_q$ with $q \in [-\infty, 1]$ and any static policy $\xi$, $w_q(E[B_T]) - E[w_q(B_T)] = O(\sqrt{T})$.

Now that we have established an upper bound on the regret of fluid policies, we proceed to seek potential tightening of the bound and to present the two different performances they have in distinct regimes of the fairness spectrum (Figure 2), $q \in (-\infty, 1]$ and $q = -\infty$. For $q \in (-\infty, 1]$, the upper bound is in fact very loose, as fluid policies are powerful enough to attain an $O(1)$ regret. In particular, the regret is exactly zero under the utilitarian welfare ($q = 1$). The proof can be found in Appendix EC.1.3.

**Theorem 1.** Fix $P \in \mathcal{P}$. Under Hölder-mean welfare $w_q$ with $q \in (-\infty, 1]$, a fluid policy $F$ attains a regret $R_T(F) = O(1)$ uniformly bounded over time horizon length $T$ solely dependent on $P$.

Theorem 1 shows the simple fluid static policy $F$ can attain an impressive $O(1)$ regret. We next provide the reader with some intuition for this result. A key property of the Hölder-mean welfare $w_q$ for $q \in (-\infty, 1]$ is its local strong smoothness around the expected utilities. More specifically, the time-average welfare $w_q(B_T/T)$ can be locally approximated with a quadratic form around the expected utilities $w_q(E[B_T]/T)$, i.e.,

$$w_q\left(\frac{B_T}{T}\right) - w_q\left(\frac{E[B_T]}{T}\right) = \nabla w_q\left(\frac{E[B_T]}{T}\right)^T (B_T - E[B_T]) - O\left(\left\|\frac{B_T - E[B_T]}{T}\right\|^2\right).$$

In the approximate quadratic form, the first-order term has zero mean by definition, and the second-order term is $O_p(T^{-1})$ by the Central Limit Theorem, as agent $i$’s utility $B^i_T = \sum_{t \in [T]} N_t \beta^i_t \xi^i_t$ is a random walk given any static policy $\xi$. Note the second-order term is negative because the welfare metric $w$ is concave. Because for some welfare metrics (such as the Nash social welfare) strong smoothness only holds locally, a technical step in the proof revolves around showing that utilities fall close their expected values with high probabilities. This implies $E[w(B_T)] - w(E[B_T]) = O(1)$, and the theorem statement follows as a result.
However, under the egalitarian welfare $w_{-\infty}$, fluid policies do not perform as well. We show that a regret of order $\sqrt{T}$ given in Proposition 1 is essentially unimprovable for fluid policies under the egalitarian welfare $w_{-\infty}$ by providing a simple example in which the fluid policy does incur an $\Omega(\sqrt{T})$ regret. See Appendix [EC.1.4] for the proof.

Lemma 4. There exists $P \in \mathcal{P}$ such that $\mathcal{R}_T(F) = \Omega(\sqrt{T})$ under the egalitarian welfare $w_{-\infty}$.

Whereas the simple fluid policies attain an $O(1)$ optimality gap for Hölder-mean welfares $w_q$ with $q \in (-\infty, 1]$, they fail to attain bounded regret for the egalitarian welfare $w_{-\infty}$. In contrast to the discussion above for smooth Hölder-mean welfare metrics (i.e., $q \in (-\infty, 1]$), the egalitarian welfare $\min_i B_i T$ is marked by its lack of smoothness at singular points where agents have equal cumulative utilities ($B_1 T = \cdots = B_n T$). Figure 3 illustrates the different behavior infinitesimal perturbations around expected utilities $E[B_T]$ may have under smooth versus non-smooth welfare metrics. In the former, the welfare metric is locally linear, so first-order welfare changes resulting from perturbations cancel out in expectation leaving only the second-order changes; in the latter, however, the welfare metric is not differentiable, so first-order welfare changes negatively impact performance. Unfortunately, the expected utility vector always lies on a singular point, i.e., the vertex of some iso-welfare hyper-surface, because the fluid policy always generate equal expected utilities across agents (see Lemma [EC.8]). The fluid policy is then doomed to a first-order welfare loss in expectation and hence a regret on the order of $\Theta(\sqrt{T})$.

Therefore, the egalitarian welfare warrants a more sophisticated approach to narrowing the regret. In the next section, we propose a computationally efficient policy based on fluid policies and show it attains $O(1)$ regret under the egalitarian welfare.

4. A policy with uniformly bounded regret under the egalitarian welfare

In this section, we focus on the egalitarian welfare $w_{-\infty}$ given by $w_{-\infty}(B_T) = \min_{i \in [n]} B_i T$. In light of fluid policies, which attain a regret of order $\Omega(\sqrt{T})$ in the worst case, we propose a collection of policies called Backward Infrequent Re-solving with Thresholding (BIRRT), detailed in
Algorithm inspired by previous fluid-based re-solving policies proposed in the literature, especially by Reiman and Wang (2008) and Bumpensanti and Wang (2020). The BIRT policies are indexed by a hyper-parameter $\eta$ and are denoted by $\text{BIRT}_\eta$, allowing for flexibility in refining the performance through control of the re-solving frequency and corresponding thresholds. The design of BIRT is characterized by the following three key elements: (i) the fluid problem is periodically re-solved after updating the deterministic surrogate quantities with their realizations so far, and the resulting policy is static between consecutive re-solving epochs; (ii) re-solving occurs only $O(\log \log T)$ times, based on a schedule designed such that the remaining schedule starting from each re-solving epoch is similar; and (iii) upon each solving and re-solving epoch, static policies are adjusted by a thresholding rule.

**Updated fluid problem.** At each re-solving epoch, by updating forecast utilities with realizations observed so far, the BIRT policies react to the stochastic deviations from the expected path initially predicted by the fluid problem. More formally, with agent $i$’s cumulative utility $B^i_{t} = \sum_{s \in [t]} b^i_s x^i_s$ by time $t \in [T]$, where $x^i_s$ is the proportion of the $s$-th resource that is allocated to agent $i$, the updated fluid problem at time $t$ is to

$$
\text{maximize} \left\{ w \left( B^i_t + (T - t) \sum_{\ell \in [L]} p^i_\ell x^i_\ell \right) : \xi \in \Delta^L_n \right\},
$$

(5)
Algorithm 2: Backward Infrequent Re-solving with Thresholding (BIRT₂)

**Input:** welfare metric \(w : \mathbb{R}^n \to \mathbb{R}\), time horizon length \(T \in \mathbb{N}\), hyper-parameter \(\eta \in (1, \frac{4}{3})\).

**Initialize:** set \(B^i_0 \leftarrow 0\) for \(i \in [n]\);

set \(K \leftarrow \lceil \log \log T / \log \eta \rceil\);

set \(t^*_0 \leftarrow 0\) and \(t^*_k \leftarrow T - \lfloor \exp(\eta^{K-k}) \rfloor\) for \(k = 1, \ldots, K\);

set \(\gamma_k \leftarrow (T - t^*_{k+1})/2n^2(T - t^*_k)\) for \(k = 0, \ldots, K-1\) and \(\gamma_K \leftarrow 0\);

for \(k = 0, 1, \ldots, K\) do

\[ \text{solve the updated fluid problem with updated welfare} \]

\[ \xi^F \in \arg \max \left\{ w \left( B^i_{t^*_k} + (T - t^*_k) \sum_{\ell \in [L]} p_{\ell i} \beta_\ell * \xi_\ell \right) : \xi \in \Delta^k_n \right\}; \]

for \(\ell = 1, 2, \ldots, L\) do // threshold the static solution

pick some agent \(j \in \arg \max_i (\xi^F)_i^j; \)

for \(i \neq j\) do

\[ \text{set} (\xi^k)_i^j \leftarrow (\xi^F)_i^j \mathbf{1}\{ (\xi^F)_i^j \geq \gamma_k \}; \]

\[ \text{set} (\xi^k)_i^j \leftarrow 1 - \sum_{i \neq j} (\xi^k)_i^j; \]

for \(t = t^*_k + 1, \ldots, t^*_{k+1}\) do // implement the thresholded static policy

observe incoming supply with utility vector \(b_t\);

for \(\ell = 1, 2, \ldots, L\) do

if \(b_t = \beta_\ell\) then

allocate \(x_t \leftarrow (\xi^k)_i^j; \)

update welfare \(B^i_t \leftarrow B^i_{t-1} + b_t x_t^i\) for every agent \(i \in [n]\);

**Output:** allocations \((x_t : t = 1, 2, \ldots, T)\) and egalitarian welfare \(\min_{i \in [n]} B^i_T\).

where \(\mathbb{E}[B^i_T | B^i_t] = B^i_t + (T - t) \sum_{\ell \in [L]} p_{\ell i} \beta_\ell \xi^F_\ell\) is the expected terminal utility of agent \(i\) under static policy \(\xi\) conditional on allocations by time \(t\).
Re-solving schedule ($\eta$). We now define the re-solving schedule, illustrated in Figure 4. Given a fixed time horizon length $T$ and a hyper-parameter $\eta \in (1, 4/3)$, we set the number of re-solving epochs as $K = \lceil \log \log T / \log \eta \rceil$. The re-solving epochs, after the initial solving epoch $t^*_0 = 0$, are given by $t^*_k = T - \lfloor \exp(\eta^{K-k}) \rfloor$ for $k \in [K]$. Remarkably, re-solving only $O(\log \log T)$ times is sufficient for a similar performance to that of “frequent” re-solving every period (see empirical evidence in Section 5.2).

An important structural feature of the re-solving schedule is its self-similarity, i.e., looking into the future from any re-solving epoch gives a similar schedule. More specifically, for a re-solving epoch $\tau$ periods from the end, the next re-solving epoch is approximately $\tau^{1/\eta}$ periods from the end. We detail in Section 4.1 how this greatly facilitates BIRT’s performance analysis.

The re-solving frequency increases over time, and most re-solving epochs are concentrated near the end of the horizon, though shown spread out in Figure 4 for illustration. Resolving more frequently near the end of the horizon is beneficial to the central planner because stochastic deviations from the predictions of the fluid policy are larger later on. In addition to incorporating more historical information, decisions closer to the end are more consequential due to the limited time left to make corrections. Indeed, the regret analysis in EC.1.5 will show that almost all regret is incurred near the end.

Thresholding Rule ($\gamma$). Given a static policy $\xi \in \Delta^L_n$ and a threshold $\gamma \in (0, 1/n)$, we define its corresponding thresholded static policy $\tilde{\xi}$ by implementing the following procedure for each $\ell \in [L]$.

After thresholding we will arrive at a static policy $\tilde{\xi}$ such that $\tilde{\xi}_i^\ell \in \{0\} \cup [\gamma, 1]$ for all $i$ and $\ell$.

1. Pick an agent that is to receive the largest type-$\ell$ allocation $j \in \arg \max_{\ell} \xi^i$. 
2. For other agents $i \neq j$, implement allocations only if they exceed the threshold $\gamma$ and withhold the rest, i.e., set $\tilde{\xi}_l^i \leftarrow \xi_l^i 1\{\xi_l^i \geq \gamma\}$.

3. Allocate the remaining resource to agent $j$, i.e., set $\tilde{\xi}_l^j \leftarrow 1 - \sum_{i \neq j} \tilde{\xi}_l^i$.

Thresholding is a defining characteristic of the $\text{BIRT}$ policies and, as shown in the experiments in Section 5.2, it is crucial to obtain uniform regret guarantees. We provide some intuition about the need for thresholding in Section 4.2.

### 4.1. Performance analysis

In this section, we first present the main result on a uniform performance guarantee of the $\text{BIRT}$ policies and then provide intuition on the regret analysis. A full proof is provided in Appendix EC.1.5.

**Theorem 2.** Fix any $P \in \mathcal{P}$ and $\eta \in (1, 4/3)$. Then, under the egalitarian welfare metric,

$$R_T(\text{BIRT}_\eta) = O(1).$$

The constant factor is determined solely by $P$ and $\eta$.

**Remark.** More generally, uniformly bounded regret can be attained by the $\text{BIRT}_\eta$ policy, even if we do not fix the distribution $P$ but rather allow for some dependency on the time horizon length $T$. Specifically, if $\min_t p_t = \Omega(T^{-1/2 + \varepsilon})$ for some $\varepsilon \in (0, 1/2]$, then $R_T(\text{BIRT}_\eta) = O(1)$ for choices of $\eta \in (1, (1 - \varepsilon/2)^{-1})$. See Theorem [EC.1] in Appendix EC.2 for a formal exposition.

The implication of Theorem 2 is twofold. The regret of non-anticipative policies against the hindsight optimum can be attributed to either an intrinsic lack of clairvoyance or suboptimal policy design. The uniform bound demonstrates that our $\text{BIRT}$ policies are near-optimal in both dimensions. On the one hand, their low regret gives credit to the policy design, in addition to the computational efficiency. On the other hand, with the non-anticipative $\text{BIRT}$ attaining performance almost comparable to a clairvoyant, Theorem 2 showcases that the central planner is able to greatly benefit from distributional knowledge.

We next give an outline of the proof, inspired by the analysis in Bumpensanti and Wang (2020), hinging on the structure of $\text{BIRT}$. In particular, the self-similar re-solving schedule warrants a serial
regret analysis. More specifically, the regret of BIRT can be decomposed into a series of expected egalitarian welfare losses against the hindsight optimum, each attributed to one re-solving epoch in the schedule. To this end, we introduce a series of auxiliary policies $\text{OPT}^k$ for $k = 0, 1, \ldots, K$, where $\text{OPT}^k$ acts exactly the same as BIRT does until $t_k^*$ and allocates in the hindsight optimal way afterwards, i.e., it picks the optimal allocation for the remaining time periods given the decisions up to time $t_k^*$. With an abuse of notation, we also denote by $\text{OPT}^k$ the egalitarian welfare generated by the $\text{OPT}^k$ policy. More precisely,

$$\text{OPT}^k := \max \left\{ w \left( B^*_i t + \sum_{t = t_k^*+1}^{T} b_t \cdot x_t \right) : x_t \in \Delta_n, t = t_k^* + 1, \ldots, T \right\},$$

(6)

where $B^*_i t$ is agent $i$’s welfare by $t$ under the BIRT policy. In particular, $\text{OPT} = \text{OPT}^0$. Notice, however, the auxiliary $\text{OPT}^k$ policy is not an online policy, as it is anticipative starting from $t_k^* + 1$.

The auxiliary policies are introduced to quantify the egalitarian welfare loss due to each re-solving epoch. For example, comparing $\text{OPT}$ and $\text{OPT}^1$ would give us that in the first epoch, as $\text{OPT}^1$ acts optimally with clairvoyance after $t_1^*$. Hence, we can decompose the regret of BIRT into

$$R_T(\text{BIRT}) = \mathbb{E} [\text{OPT} - \text{ALG}(\text{BIRT})] = \sum_{k=1}^{K} \mathbb{E} \left[ \text{OPT}^{k-1} - \text{OPT}^k \right] + \mathbb{E} \left[ \text{OPT}^K - \text{ALG}(\text{BIRT}) \right].$$

The telescopic decomposition is useful for two reasons. First, since the problems $\text{OPT}^{k-1}$ and $\text{OPT}^k$ share the same decisions up to time $t_{k-1}^*$ and the re-solving schedule is self-similar, their difference in egalitarian welfare $\text{OPT}^{k-1} - \text{OPT}^k$ is isomorphic to $\text{OPT} - \text{OPT}^1$, with appropriately chosen initial welfare and horizon lengths. Second, we can analyze each term $\mathbb{E} [\text{OPT}^{k-1} - \text{OPT}^k]$ without considering inter-epoch interactions, i.e., consequences of previous allocations on subsequent periods. In other words, analyses of disjoint epochs are decoupled from one another. This follows because we can provide guarantees for each term that are uniform over the agents’ initial utilities.

The decomposition warrants an inductive argument to show the main theorem using the following result on one-epoch expected egalitarian welfare loss. For the purpose of providing a uniform framework for analyzing each re-solving epoch, we assume the agents start with initial utilities $B_0 \in \mathbb{R}_n^+$ that may be positive, so that the analysis can be directly deployed to subsequent epochs.
Proposition 2. Fix an arrival distribution $P \in \mathcal{P}$, initial utilities $B_0 \in \mathbb{R}_+^n$ and time horizon length $T > 1$. Suppose the first re-solving occurs at some $t_1 \in [T]$, and the initial threshold is set as $\gamma_0 = (T-t_1)/2n^2T$, then for a constant $C > 0$ determined by $P$ and independent of $T$, $t_1$ and $B_0$, 

$$
E[\text{OPT} - \text{OPT}^1|B_0] \leq 3Lt_1 \exp\left(-C\frac{(T-t_1)^4}{T^3}\right).
$$

The complete proof is postponed to Appendix EC.1.6. Its main idea is to focus on the coupling event $\{\text{OPT} = \text{OPT}^1\}$, where the egalitarian welfare loss is zero. On the complementary event $\{\text{OPT} \neq \text{OPT}^1\}$, we can control the difference using the almost sure uniform bound $\text{OPT} - \text{OPT}^1 \leq t_1$ (Lemma EC.1), which follows because any loss the clairvoyant $\text{OPT}^1$ can incur against $\text{OPT}$ must only stem from the suboptimal allocations by the static policy during the first $t_1$ periods. The most complex step of the proof involves showing that $\{\text{OPT} = \text{OPT}^1\}$ is a high probability event. To do so, we argue that $\text{OPT}$ and $\text{OPT}^1$ generate the same egalitarian welfare if $\text{OPT}^1$ manages to replicate the exact aggregate allocations of each type to each agent made by $\text{OPT}$; we invoke sensitivity analysis results for linear programming to provide sufficient conditions in terms of the arrival process; we then prove that these sufficient conditions hold with high probability using concentration inequalities. Thresholding the fluid policy is critical to proving that the concentration inequalities hold uniformly over the initial agents’ utilities as, otherwise, BIRT would likely fall into the trap of not being able to replicate hindsight optimal allocations due to irrevocability of allocations.

4.2. Degeneracy and the need for thresholding

The fair allocation problem under the egalitarian welfare can be cast as a linear program by writing the objective in the epigraph form:

$$
\begin{align*}
\max_{u, \xi} & \quad u \\
\text{s.t.} & \quad u \leq \sum_{\ell \in [L]} \mathbb{E}[N_\ell] \beta^i \xi^i_{\ell}, \quad \forall i \in [n], \\
& \quad \sum_{i \in [n]} \xi^i_{\ell} \leq 1, \quad \forall \ell \in [L], \\
& \quad \xi^i_{\ell} \geq 0, \quad \forall \ell \in [L], i \in [n].
\end{align*}
$$

(8)
We refer to the fluid problem as *degenerate* if any of its optimal solutions is degenerate as defined for linear programs. More precisely, a fluid policy $F$ is degenerate if more constraints are active than the dimension of optimization variables, i.e.,

$$\left| \left\{ i \in [n] : \text{FLU} = \sum_{\ell \in [L]} E[N_{\ell}] \beta_{i,\ell}^i \right\} \right| + \left| \left\{ \ell \in [L] : \sum_{i \in [n]} (\xi_F)_{i,\ell} = 1 \right\} \right| + \left| \left\{ (\ell, i) : (\xi_F)_{i,\ell} = 0 \right\} \right| > nL + 1.$$

When the fluid problem (8) is non-degenerate, the same basis is optimal both in the hindsight problem (2) when the number of arrivals of each type $N_{\ell}$ is close to its expected value $T_{p_{\ell}}$ and in the updated fluid problem (5) when the cumulative agents' utilities $B_i^t$ follow likely trajectories. Therefore, with high probability, the hindsight problem and the updated fluid problems have the same optimal basis, which leads to similar allocations. When the problem is non-degenerate, a vanilla resolving policy without thresholding can achieve constant regret.

Thresholding plays a key role when the fluid problem is degenerate (or close to degenerate). In this case, slight perturbations in the number of arrivals of each type or the agents' utilities can lead to changes in the optimal basis of the linear program. A vanilla resolving policy without thresholding could introduce welfare losses by choosing bases different from that of the hindsight problem. Intuitively, by adjusting small allocation components to zero, thresholding essentially forces adopting a degenerate basic solution to the fluid problem corresponding to multiple bases. Thresholding postpones selecting any basis by placing BIRT on multiple bases that are sufficiently close by so that past suboptimal allocation decisions could be made up for later once the final choice of basis is revealed. In a word, the gist of thresholding is to *postpone choosing any of neighboring bases until subsequent arrivals reveal more information.*

An alternative way to understand thresholding is through our theoretical regret analysis in Section 4.1. The intuition behind thresholding is that BIRT aims to recover the hindsight optimum by the end of the horizon; in particular and more ambitiously, it aspires to replicate the exact amount of resources of each type $\ell$ allocated to each agent $i$ by the hindsight optimal policy. Though re-solving helps approximate these hindsight optimal allocations, exact replication is hindered by stochasticity, especially for $(\ell, i)$ where the fluid policy yields allocations $(\xi^F)_{i,\ell}$ close to zero. As a
result of stochasticity, the hindsight optimal policy may possibly end up allocating none of type-\(\ell\) resources to agent \(i\), in which case the replication goal fails completely as allocations are irrevocable. Hence, it is best to altogether adjust these allocations to zero, leaving more leeway for better informed allocations in the future.

Moreover, thresholding becomes more aggressive with larger thresholds over time. While thresholding allows to postpone allocation decisions that are close to zero or one, it introduces artificial adjustments that deviate from the fluid “optimal” path. To minimize such deviation from expected paths, the threshold should be smaller if the waiting time to the next re-solving epoch is large in relative terms, as later re-solving poses a larger risk of excessive deviation. Since the proportion of time left to the next re-solving epoch \((t_{k+1}^* - t_k^*)/(T - t_k^*)\) decreases over \(k = 0, \ldots, K - 1\), thresholds increase accordingly, leading to more aggressive thresholding toward the end.

4.3. Technical novelty

Before ending this section, we would like to comment on some key distinctions between our problem and analyses from those in Bumpensanti and Wang (2020), whose work in network revenue management inspired the BIRT policy.

The most salient feature of our problem is that its objective is non-linear and hence non-separable over time. This renders inapplicable a direct decomposition of the time horizon into several, disjoint re-solving epochs and plugging in single-epoch regret guarantees (assuming memorylessness of the allocations). Rather, in our problem, each re-solving epoch starts with the random utilities garnered from stochastic arrivals in previous epochs, and these initial conditions have to be taken into account in epoch-wise analyses.

In addition, in the dynamic fair allocation problem, the central planner needs to assign resources across multiple agents, i.e., allocation decisions lie in the simplex \(\Delta_n\) across the \(n\) agents. In the network revenue management problem, the decision maker needs to decide whether to fulfill demand requests, i.e., allocation decisions are binary accept-reject. This discrepancy complicates
the thresholding procedure as thresholding one agent’s assignment affects the decisions of all other agents.

In analyzing the performance, Bumpensanti and Wang (2020) invoke existing results by Reiman and Wang (2008) for network revenue management problems, whereas we provide a new and simpler stand-alone analysis that exploits the structure of our problem. Our analysis also facilitates a more fundamental understanding of the impact of stochastic arrivals on the optimal solution through a new sensitivity analysis and of the coupling between the hindsight optimal solution and auxiliary policies.

5. Experiments

In this section, we present numerical results from our simulation experiments to illustrate the performances of the fluid and BiRT policies. The first experiment evaluates these policies on several randomized arrival distributions under the egalitarian welfare, the harmonic-mean welfare and the Nash social welfare. The second experiment focuses on the special case of the egalitarian welfare and exposes the impact of degeneracy on re-solving heuristics. It is built upon two artificially designed arrival distributions, whose fluid problems are degenerate and nondegenerate respectively.

5.1. Randomized experiment

In the first experiment, we illustrate the performances of the F and BiRT policies under the egalitarian welfare, the harmonic-mean welfare and the Nash social welfare, using a number of randomized problem instances. More specifically, we consider a problem setup of \( n = 4 \) agents and \( L = 5 \) types of resources. We simulate probabilities of the resource types uniformly at random from the simplex, i.e., \( (p_\ell : \ell \in [L]) \sim \text{Unif}(\Delta_L) \), and the marginal utilities independently and identically from a parameterized Beta distribution, i.e., \( \beta^i_\ell \text{ iid } \text{Beta}(\cdot, \cdot) \) across \( \ell, i \). In particular, we make the two choices for the marginal utilities:

- \( \beta^i_\ell \sim \text{Beta}(0.5, 0.5) \), under which values are concentrated around the two extremes of 0 and 1, or in other words, the agents have dichotomous valuations for the resources;
• $\beta_i^\ell \sim \text{Beta}(2.0, 2.0)$, under which values are concentrated near the middle at 0.5, or in other words, the agents have similar valuations for the resources.

In each case, we simulate 30 instances of arrival distributions and compute the average relative regrets over the randomized arrival distributions. Note that relative regrets are used to normalize the problems for a reasonable comparison regardless of welfare dimensions. Results are shown in Figure 5. The meta-distributions for generating arrival distributions are distinguished by marks, and the policies are distinguished by colors.

We summarize our findings under the three welfare metrics.

• Figure 5a shows that under the egalitarian welfare ($q = -\infty$), the F policy attains a relative regret of $\rho_T(F) = \Theta(T^{-1/2})$, translating to a regret of $R_T(F) = \Theta(\sqrt{T})$; the BIRT policy attains a relative regret of $\rho_T(BIRT) = \Theta(T^{-1})$, translating to a uniformly bounded regret of $R_T(BIRT) = \Theta(1)$.

• Figure 5b shows that under the harmonic-mean welfare ($q = -1$), the F policy attains a relative regret of $\rho_T(F) = \Theta(T^{-1})$, translating to a regret of $R_T(F) = \Theta(1)$; the BIRT policy attains a relative regret of $\rho_T(BIRT) = o(T^{-1})$, translating to a vanishing regret of $R_T(BIRT) = o(1)$.

• Figure 5c shows that under the Nash social welfare ($q = 0$), the F policy attains a relative regret of $\rho_T(F) = \Theta(T^{-1})$, translating to a regret of $R_T(F) = \Theta(1)$; the BIRT policy attains a relative regret of $\rho_T(BIRT) = o(T^{-1})$, translating to a vanishing regret of $R_T(BIRT) = o(1)$.

These empirical findings from the randomized experiments exactly match our theoretical guarantees. Interestingly, we notice the BIRT policy attains a resounding $o(1)$ vanishing regret under the harmonic-mean and Nash social welfare metrics, an even more impressive improvement upon the $\Theta(1)$ regret attained by the static F policy. The precise mechanism to attaining the vanishing regret is unknown to us, but the empirical findings do highlight the excellent performance that BIRT has under all welfare metrics.

5.2. Special experiment under the egalitarian welfare

In the second experiment, we zoom in and focus on the egalitarian welfare. We aim to corroborate our claims on the performances of the fluid and BIRT policies under the egalitarian welfare and,
Figure 5  Average relative regret $\mathcal{P}(\pi)$ over randomized arrival distributions, under the egalitarian welfare ($q = -\infty$), the harmonic-mean welfare ($q = -1$) and the Nash social welfare ($q = 0$), plotted against the time horizon length $T$ on a double logarithmic scale

moreover, to understand what driving forces are behind the outstanding performance of BIRT. To be more precise, we are including the following heuristic policies for comparison.

1. $F$: the fluid static policy based on the initial fluid problem (Algorithm\[1\]).

2. BIRT: the Backward Infrequent Re-solving with Thresholding policies (Algorithm\[2\]).
3. **BIR**: the Backward Infrequent Re-solving policies (Algorithm 3 in Appendix EC.4), which re-solve the fluid problems on the same schedule as **BIRT** does but without thresholding (i.e., the updated fluid policy $\xi^F$ at re-solving epoch $t^*_k$ is directly used as $\xi^k$ for implementation during the current epoch).

4. **FR**: the Frequent Re-solving policy (Algorithm 4 in Appendix EC.4), which re-solves the fluid problems in every period before making allocation decisions.

We include the **BIR** policy (Algorithm 3) to illustrate the crucial role that thresholding is playing in improving performance and the **FR** policy (Algorithm 4) to evaluate whether performance could be improved by substituting thresholding with more frequent re-solving.

In addition to inter-policy comparison, we would like to investigate empirically what impact a degenerate deterministic problem has on the performance of re-solving heuristics, as widely discussed in previous literature.

To illustrate the impact of degeneracy, we consider a simple problem setup with $n = 2$ agents and $L = 2$ types of resources where the marginal utilities are $\beta_1 = (1, 1/2)$ and $\beta_2 = (1/2, 1)$. We then consider two arrival distributions $(p_1, p_2) = (1/2, 1/2)$ and $(p_1, p_2) = (2/5, 3/5)$. The first case (detailed in Appendix EC.1.4) results in a degenerate fluid problem, as a total of $6 > 5$ constraints are active at the fluid policy $\xi^F$ with $(\xi^F)_1 = (1, 0)$ and $(\xi^F)_2 = (0, 1)$, corresponding to the two bases $(u, \xi^1_1, \xi^2_1, \xi^1_2)$ and $(u, \xi^1_1, \xi^2_2, \xi^2_2)$. Therefore, slight perturbations on the number of arrivals can lead to optimal hindsight allocations of the form $(\xi)_1 = (1, 0)$ and $(\xi)_2 = (\epsilon, 1 - \epsilon)$ or $(\xi)_1 = (1 - \epsilon, \epsilon)$ and $(\xi)_2 = (0, 1)$ for $\epsilon \geq 0$, and, without thresholding, the resolving policy might allocate resources to the wrong agent (e.g., allocating resource one to agent two when the hindsight optimal basis is the first one). In the second case, the original fluid problem is nondegenerate as a total of $5$ constraints are active at the fluid policy $\xi^F$ with $(\xi^F)_1 = (1, 0)$ and $(\xi^F)_2 = (2/9, 7/9)$, corresponding to the unique basis $(u, \xi^1_1, \xi^2_1, \xi^2_2)$. Therefore, with high probability, an optimal hindsight allocation is of the form $(\xi)_1 = (1, 0)$ and $(\xi)_2 = (2/9 \pm \epsilon, 7/9 \mp \epsilon)$ and it is less likely to allocate resources to the wrong agent.

In both cases, the same hyper-parameter $\eta = 1.05$ is chosen for the **BIRT** and **BIR** policies. Regrets of respective policies are shown in Figure 6.
First, we observe from Figure 6 that in both cases BIRT’s regret converges, corroborating our claim (Theorem 2) that it attains $O(1)$ regret. Equipped with distributional knowledge, the central planner is shown to be able to improve the performance upon the fluid static policy by re-solving infrequently (BIR policy), either by improving the factor of $\sqrt{T}$ in the regret in the degenerate case, or by attaining a constant regret in the nondegenerate case. Interestingly, in both cases, further increasing the re-solving frequency does not lead to performance improvement; in fact, the BIR and FR policies attain similar regrets. In other words, carefully designing the re-solving schedule suffices to guarantee good performance while drastically reducing computational complexity. The most significant difference between the two cases is that the BIR and FR policies attain $\Theta(\sqrt{T})$ regret in the degenerate case, whereas they can attain $O(1)$ regret in the nondegenerate case, indicating their vulnerability against degeneracy. By contrast, the more robust BIRT policy steadily attains a $O(1)$ regret in both cases, demonstrating that the thresholding rule is crucial for controlling the excessive variation in allocations and obtain remarkable performance even in the nondegenerate case.

6. Conclusion

In this work we study dynamic policies for allocating $T$ divisible items arriving sequentially to a fixed set of $n$ agents. We consider a collection of Hölder-mean welfare metrics, including the
Nash social welfare and the egalitarian welfare, and evaluate policies based on the expected welfare generated, which entails addressing the trade-off between efficiency and fairness. We show the classical fluid policy based on certainty equivalents incurs a $\Theta(\sqrt{T})$ regret against the expected hindsight optimal egalitarian welfare and an $O(1)$ regret under all other welfare metrics. For the egalitarian welfare, we propose the Backward Infrequent Re-solving with Thresholding (BIRT) policies, which incur an $O(1)$ regret, uniformly bounded over time horizon length $T$.

It would be interesting to extend the work to problems defined on an infinite time horizon, or with infinitely many arrival types. To allow for flexibility in the arrival model, an open question is to design dynamic allocation policies that are robust against inexact solutions to the fluid problems and distribution mis-specification in the IID case (inaccurate utilities), and that work well when arrivals are inter-temporally dependent (e.g., autocorrelated, evolving under a Markov chain model).

References

Abdulkadiroğlu A, Che YK, Pathak PA, Roth AE, Tercieux O (2017) Minimizing justified envy in school choice: the design of new orleans’ oneapp. Technical report, National Bureau of Economic Research.

Abdulkadiroğlu A, Sönmez T (2003) School choice: A mechanism design approach. American economic review 93(3):729–747.

Agarwal N, Ashlagi I, Rees MA, Somaini PJ, Waldinger DC, et al. (2019) An empirical framework for sequential assignment: The allocation of deceased donor kidneys (National Bureau of Economic Research).

Agrawal S, Devanur NR (2014) Fast algorithms for online stochastic convex programming. Proceedings of the twenty-sixth annual ACM-SIAM symposium on Discrete algorithms, 1405–1424 (SIAM).

Ahmed F, Dickerson JP, Fuge M (2017) Diverse weighted bipartite b-matching. arXiv preprint arXiv:1702.07134.

Aleksandrov M, Walsh T (2020) Online fair division: A survey. Proceedings of the AAAI Conference on Artificial Intelligence, volume 34, 13557–13562.

Aleksandrov MD, Aziz H, Gaspers S, Walsh T (2015) Online fair division: Analysing a food bank problem. Twenty-Fourth International Joint Conference on Artificial Intelligence.
Alkaabneh F, Diabat A, Gao HO (2021) A unified framework for efficient, effective, and fair resource allocation by food banks using an approximate dynamic programming approach. *Omega* 100:102300.

Arlotto A, Gurvich I (2019) Uniformly bounded regret in the multisecretary problem. *Stochastic Systems* 9(3):231–260.

Arnosti N, Beck M, Milgrom P (2016) Adverse selection and auction design for internet display advertising. *American Economic Review* 106(10):2852–66.

Arnosti N, Shi P (2020) Design of lotteries and wait-lists for affordable housing allocation. *Management Science* 66(6):2291–2307.

Arora S, Hazan E, Kale S (2012) The multiplicative weights update method: a meta-algorithm and applications. *Theory of computing* 8(1):121–164.

Asadpour A, Saberi A (2010) An approximation algorithm for max-min fair allocation of indivisible goods. *SIAM Journal on Computing* 39(7):2970–2989.

Azar Y (1998) On-line load balancing. *Online algorithms* 178–195.

Azar Y, Epstein L (1998) On-line machine covering. *Journal of Scheduling* 1(2):67–77.

Azar Y, Naor J, Rom R (1995) The competitiveness of on-line assignments. *Journal of Algorithms* 18(2):221–237.

Balseiro S, Besbes O, Pizarro D (2021a) Survey of dynamic resource constrained reward collection problems: Unified model and analysis. *Available at SSRN 3963265*.

Balseiro S, Lu H, Mirrokni V (2021b) Regularized online allocation problems: Fairness and beyond. *International Conference on Machine Learning*, 630–639 (PMLR).

Banerjee S, Gkatzelis V, Gorokh A, Jin B (2022a) Online nash social welfare maximization with predictions. *Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 1–19, URL [http://dx.doi.org/10.1137/1.9781611977073.1](http://dx.doi.org/10.1137/1.9781611977073.1)

Banerjee S, Gkatzelis V, Hossain S, Jin B, Micha E, Shah N (2022b) Proportionally fair online allocation of public goods with predictions. *arXiv preprint arXiv:2209.15305*.

Bansal N, Sviridenko M (2006) The santa claus problem. *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, 31–40.
Bateni M, Chen Y, Ciocan DF, Mirrokni V (2022) Fair resource allocation in a volatile marketplace. *Operations Research* 70(1):288–308.

BEN-ARI M (1998) How to solve the santa claus problem. *Concurrency: Practice and Experience* 10(6):485–496.

Benade G, Kazachkov AM, Procaccia AD, Psomas CA (2018) How to make envy vanish over time. *Proceedings of the 2018 ACM Conference on Economics and Computation*, 593–610.

Bertsekas D (2012) *Dynamic programming and optimal control: Volume I*, volume 1 (Athena scientific).

Bertsimas D, Farias VF, Trichakis N (2011) The price of fairness. *Operations research* 59(1):17–31.

Bertsimas D, Farias VF, Trichakis N (2012) On the efficiency-fairness trade-off. *Management Science* 58(12):2234–2250.

Bezáková I, Dani V (2005) Allocating indivisible goods. *ACM SIGecom Exchanges* 5(3):11–18.

Brandt F, Conitzer V, Endriss U, Lang J, Procaccia AD (2016) *Handbook of computational social choice* (Cambridge University Press).

Bray RL (2019) Logarithmic regret in multisecretary and online linear programming problems with continuous valuations. *arXiv e-prints* arXiv–1912.

Bumpensanti P, Wang H (2020) A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science* 66(7):2993–3009.

Caragiannis I, Kurokawa D, Moulin H, Procaccia AD, Shah N, Wang J (2019) The unreasonable fairness of maximum nash welfare. *ACM Trans. Econ. Comput.* 7(3), ISSN 2167-8375, URL [http://dx.doi.org/10.1145/3355902](http://dx.doi.org/10.1145/3355902)

Cayci S, Gupta S, Eryilmaz A (2020) Group-fair online allocation in continuous time. *Advances in Neural Information Processing Systems* 33:13750–13761.

Chakrabarty D, Chuzhoy J, Khanna S (2009) On allocating goods to maximize fairness. *2009 50th Annual IEEE Symposium on Foundations of Computer Science*, 107–116 (IEEE).

Cheung WC, Lyu G, Teo CP, Wang H (2020) Online planning with offline simulation. *Available at SSRN 3709882*.
Choi H, Mela CF, Balseiro SR, Leary A (2020) Online display advertising markets: A literature review and future directions. *Information Systems Research* 31(2):556–575.

Conitzer V, Freeman R, Shah N (2017) Fair public decision making. *Proceedings of the 2017 ACM Conference on Economics and Computation*, 629–646, EC ’17 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450345279, URL [http://dx.doi.org/10.1145/3033274.3085125](http://dx.doi.org/10.1145/3033274.3085125).

Deuermeyer BL, Friesen DK, Langston MA (1982) Scheduling to maximize the minimum processor finish time in a multiprocessor system. *SIAM Journal on Algebraic Discrete Methods* 3(2):190–196.

Devanur NR, Jain K (2012) Online matching with concave returns. *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing*, 137–144, STOC ’12 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450312455, URL [http://dx.doi.org/10.1145/2213977.2213992](http://dx.doi.org/10.1145/2213977.2213992).

Devanur NR, Jain K, Sivan B, Wilkens CA (2019) Near optimal online algorithms and fast approximation algorithms for resource allocation problems. *Journal of the ACM (JACM)* 66(1):1–41.

Dickerson JP, Sankararaman KA, Srinivasan A, Xu P (2019) Balancing relevance and diversity in online bipartite matching via submodularity. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 1877–1884.

Dinh TQ, Liang B, Quek TQ, Shin H (2020) Online resource procurement and allocation in a hybrid edge-cloud computing system. *IEEE transactions on wireless communications* 19(3):2137–2149.

Donahue K, Kleinberg J (2020) Fairness and utilization in allocating resources with uncertain demand. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 658–668.

Feige U (2008) On allocations that maximize fairness. *SODA*, volume 8, 287–293 (Citeseer).

Freeman R, Shah N, Vaish R (2020) Best of both worlds: Ex-ante and ex-post fairness in resource allocation. *Proceedings of the 21st ACM Conference on Economics and Computation*, 21–22, EC ’20 (New York, NY, USA: Association for Computing Machinery), ISBN 9781450379755, URL [http://dx.doi.org/10.1145/3391403.3399537](http://dx.doi.org/10.1145/3391403.3399537).

Freeman R, Zahedi SM, Conitzer V (2017) Fair and efficient social choice in dynamic settings. *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI-17*, 4580–4587, URL [http://dx.doi.org/10.24963/ijcai.2017/639](http://dx.doi.org/10.24963/ijcai.2017/639).
Gallego G, Van Ryzin G (1997) A multiproduct dynamic pricing problem and its applications to network yield management. *Operations research* 45(1):24–41.

Gálvez W, Soto JA, Verschae J (2020) Symmetry exploitation for online machine covering with bounded migration. *ACM Transactions on Algorithms (TALG)* 16(4):1–22.

Gollapudi S, Panigrahi D (2014) Fair allocation in online markets. *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 1179–1188.

Golovin D (2005) *Max-min fair allocation of indivisible goods* (School of Computer Science, Carnegie Mellon University).

Grigoryan A (2021) Effective, fair and equitable pandemic rationing. *Available at SSRN 3646539*.

Haeupler B, Saha B, Srinivasan A (2011) New constructive aspects of the lovász local lemma. *Journal of the ACM (JACM)* 58(6):1–28.

Hahn F, Arrow K, Intriligator M (1982) Handbook of mathematical economics.

Hardy GH, Littlewood JE, Pólya G, Pólya G, et al. (1952) *Inequalities* (Cambridge university press).

He J, Procaccia AD, Psomas C, Zeng D (2019) Achieving a fairer future by changing the past. *IJCAI’19*.

Jasin S, Kumar S (2012) A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research* 37(2):313–345.

Kaplan EH (1984) *Managing the demand for public housing*. Ph.D. thesis, Massachusetts Institute of Technology.

Kawase Y, Sumita H (2020) On the max-min fair stochastic allocation of indivisible goods. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2070–2078.

Kawase Y, Sumita H (2021) Online max-min fair allocation. *arXiv preprint arXiv:2111.07235*.

Li X, Rong Y, Zhang R, Zheng H (2021) Online advertisement allocation under customer choices and algorithmic fairness. *Available at SSRN 3538755*.

Li Y, He C, Jiang Y, Wu W, Jiang J, Zhang W, Fan H (2019) Max-min fair allocation for resources with hybrid divisibilities. *Expert Systems with Applications* 124:325–340.
Lien RW, Iravani SM, Smilowitz KR (2014) Sequential resource allocation for nonprofit operations. *Operations Research* 62(2):301–317.

Luce RD, Raiffa H (1989) *Games and decisions: Introduction and critical survey* (Courier Corporation).

Ma W, Xu P, Xu Y (2020) Group-level fairness maximization in online bipartite matching. *arXiv preprint arXiv:2011.13908*.

Manshadi V, Niazadeh R, Rodilitz S (2021) Fair dynamic rationing. *Proceedings of the 22nd ACM Conference on Economics and Computation*, 694–695.

Moulin H (2004) *Fair division and collective welfare* (MIT press).

Nanda V, Xu P, Sankararaman KA, Dickerson J, Srinivasan A (2020) Balancing the tradeoff between profit and fairness in rideshare platforms during high-demand hours. *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2210–2217.

Nash Jr JF (1950) The bargaining problem. *Econometrica: Journal of the econometric society* 155–162.

Pathak PA, Rees-Jones A, Sönmez T (2020) Immigration lottery design: engineered and coincidental consequences of h-1b reforms. Technical report, National Bureau of Economic Research.

Rawls J (2001) *Justice as fairness: A restatement* (Harvard University Press).

Rawls J (2020) A theory of justice. *A theory of justice* (Harvard university press).

Reiman MI, Wang Q (2008) An asymptotically optimal policy for a quantity-based network revenue management problem. *Mathematics of Operations Research* 33(2):257–282.

Schrijver A (1998) *Theory of linear and integer programming* (John Wiley & Sons).

Sinclair SR, Jain G, Banerjee S, Yu CL (2020) Sequential fair allocation of limited resources under stochastic demands. *arXiv preprint arXiv:2011.14382*.

Talluri K, Van Ryzin G (1998) An analysis of bid-price controls for network revenue management. *Management science* 44(11-part-1):1577–1593.

Varian HR (1974) Equity, envy, and efficiency. *Journal of Economic Theory* 9(1):63–91, ISSN 0022-0531, URL http://dx.doi.org/https://doi.org/10.1016/0022-0531(74)90075-1.
Vera A, Banerjee S (2021) The bayesian prophet: A low-regret framework for online decision making. *Management Science* 67(3):1368–1391.

Vera A, Banerjee S, Gurvich I (2021) Online allocation and pricing: Constant regret via bellman inequalities. *Operations Research* 69(3):821–840.

Watts A (2021) Fairness and efficiency in online advertising mechanisms. *Games* 12(2):36.

Woeginger GJ (1997) A polynomial-time approximation scheme for maximizing the minimum machine completion time. *Operations Research Letters* 20(4):149–154.

Wu H, Srikant R, Liu X, Jiang C (2015) Algorithms with logarithmic or sublinear regret for constrained contextual bandits. *Advances in Neural Information Processing Systems* 28.

Wu X, Li B, Gan J (2021) Budget-feasible maximum nash social welfare is almost envy-free (International Joint Conferences on Artificial Intelligence Organization).

Zeng D, Psomas A (2020) Fairness-efficiency tradeoffs in dynamic fair division. *Proceedings of the 21st ACM Conference on Economics and Computation*, 911–912.

Zhang J, Xie N, Zhang X, Li W (2018) An online auction mechanism for cloud computing resource allocation and pricing based on user evaluation and cost. *Future Generation Computer Systems* 89:286–299.
Appendices

EC.1. Proofs

EC.1.1. Proof of Lemma 2

Fix $P \in \mathcal{P}$. By Lemma EC.7, $P$-almost surely, there is a static policy $\xi^*$ that is hindsight optimal, i.e., $\text{OPT} = w(\sum_{\ell \in [L]} N_\ell \beta_\ell \ast \xi^*_\ell)$. Consider an alternate static policy $\xi \in \Delta^L_n$ given by $\xi_i^\ell := \frac{E[N_\ell(\xi^*)_i]}{E[N_\ell]}$, for all $i, \ell$. Hence,

$$E[\text{OPT}] = E\left[w\left(\sum_{\ell \in [L]} N_\ell \beta_\ell \ast \xi^*_\ell\right)\right] \leq w\left(\sum_{\ell \in [L]} \beta_\ell^i E[N_\ell(\xi^*)_\ell]\right) = w\left(\sum_{\ell \in [L]} E[N_\ell] \beta^i_\ell \xi^*_\ell\right) \leq \text{FLU},$$

where the first equality follows by definition of $\xi^*$, the first inequality by Jensen’s inequality, the second equality by definition of $\xi$, and the second inequality by definition of $\text{FLU}$.

EC.1.2. Proof of Lemma 3

Fix static policy $\xi$ and welfare metric $w \equiv w_q$—we omit the notation for convenience in the proof. Denote $p := \min_\ell p_\ell$. Clearly, we can assume $w(E[B_T]) > 0$ because the lemma is otherwise trivial. First notice

$$w(B_T) \geq w\left(E[B_T] \min_i \frac{B^i_T}{E[B^i_T]}\right) = w(E[B_T]) \min_i \frac{B^i_T}{E[B^i_T]},$$

where the inequality follows because $w$ is nonnegative and monotone, and the equality because $w$ is positively homogeneous. Hence,

$$1 - \frac{E[w(B_T)]}{w(E[B_T])} \leq E\left[\max_\ell \frac{E[B^\ell_T]}{E[B^\ell_T]} - \frac{B^\ell_T}{E[B^\ell_T]}\right] \leq E\left(\max_\ell \frac{E[B^\ell_T]}{E[B^\ell_T]} - \frac{B^\ell_T}{E[B^\ell_T]}\right)^2 \leq \sum_\ell E\left(\frac{E[B^\ell_T]}{E[B^\ell_T]} - \frac{B^\ell_T}{E[B^\ell_T]}\right)^2 \leq \sqrt{L \frac{pT}{T}},$$

where the first inequality follows because $w(E[B_T]) > 0$, the second by Jensen’s inequality, the third by nonnegativity of squares, and the last by Lemma EC.11. Since $w$ is positively homogeneous, $w(E[B_T]) = T w(\sum_{\ell \in [L]} p_\ell \beta_\ell \ast \xi_\ell)$, so

$$w(E[B_T]) - E[w(B_T)] \leq \sqrt{\frac{LT}{p}} w\left(\sum_{\ell \in [L]} p_\ell \beta_\ell \ast \xi_\ell\right) \leq \frac{1}{n} \sqrt{\frac{LT}{p}},$$

where the last inequality follows by Lemma EC.10. The proof is thereby concluded. □
EC.1.3. Proof of Theorem \[1\]

Fix static policy \( \xi \) and omit the notation for convenience. We show the following claim, to which the theorem is a corollary by invoking Lemma \[2\].

**Claim:** Under any Hölder-mean welfare metric \( w_q \) with \( q \in (-\infty, 1] \) and any static policy \( \xi \), we have \( w_q(\mathbb{E}[B_T]) - \mathbb{E}[w_q(B_T)] = O(1) \) uniformly bounded over time horizon length \( T \).

We assume \( w_q(\mathbb{E}[B_T]) > 0 \) because the theorem is trivial otherwise. For convenience, we denote \( p := \min_i p_i > 0 \) and define \( Z^i := B^i_T/\mathbb{E}[B^i_T] - 1 \) for each agent \( i \) if \( \mathbb{E}[B^i_T] > 0 \) and \( Z^i = 0 \) otherwise. Then \( \mathbb{E}[Z^i] = 0 \) and \( Z^i \in [-1, 1/p] \) a.s. because \( B^i_T \geq 0 \) and

\[
1 + Z^i = \frac{B^i_T}{\mathbb{E}[B^i_T]} = \frac{\sum T_i N_i \beta^i_{\ell, i} \xi^i_{\ell}}{\sum T_i p_i \beta^i_{\ell, i}} \leq \frac{\sum T_i \beta^i_{\ell, i}}{p} \quad \text{a.s. (EC.1)}
\]

because \( N_\ell \leq T \) for all \( \ell \in [L] \) and \( p_\ell \geq p \).

In the sequel, we show the claim separately for \( q \in (0, 1) \), \( q = 0 \) and \( q \in (-\infty, 0) \).

**Proof of Theorem \[1\] for \( q \in (0, 1) \).** First notice Lemma [EC.15] implies

\[
\frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \geq 1 + \frac{1}{q} \sum_{i \in [n]} (\mathbb{E}[B^i_T])^q [(1 + Z^i)^q - 1] \geq 1 + \frac{\sum_{i \in [n]} (\mathbb{E}[B^i_T])^q [Z^i - (q^{-1} - 1)(Z^i)^2]}{\sum_{i \in [n]} (\mathbb{E}[B^i_T])^q},
\]

so taking expectations yields

\[
1 - \frac{\mathbb{E}[w_q(B_T)]}{w_q(\mathbb{E}[B_T])} \leq (q^{-1} - 1) \frac{\sum_{i \in [n]} (\mathbb{E}[B^i_T])^q \mathbb{E}[(Z^i)^2]}{\sum_{i \in [n]} (\mathbb{E}[B^i_T])^q} \leq \frac{q^{-1} - 1}{p} T^{-1},
\]

where the first inequality follows because \( \mathbb{E}[Z^i] = 0 \), and the second by Lemma [EC.11].

**Proof of Theorem \[1\] for Nash social welfare, i.e., \( q = 0 \).** Consider the event \( E_{1/2} = \{|Z^i| \leq 1/2, \forall i\} \), on which

\[
\frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \geq 1 + \log \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} = 1 + \frac{1}{n} \sum_{i \in [n]} \log(1 + Z^i) \geq 1 + \frac{1}{n} \sum_{i \in [n]} [Z^i - (Z^i)^2],
\]
where the first inequality follows because $\exp(x) \geq 1 + x$ for all $x \in \mathbb{R}$, the equality by definition of $w$ and $Z^i$, and the second inequality by Lemma [EC.14]. We then notice

$$
\mathbb{E} \left[ \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \right] \geq \mathbb{E} \left[ \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])}; E_{1/2} \right]
$$

$$
\geq \mathbb{E} \left[ 1 + \frac{1}{n} \sum_{i \in [n]} Z^i - \frac{1}{n} \sum_{i \in [n]} (Z^i)^2; E_{1/2} \right]
$$

$$
\geq \mathbb{E} \left[ 1 + \frac{1}{n} \sum_{i \in [n]} Z^i; E_{1/2} \right] - \frac{1}{n} \sum_{i \in [n]} \mathbb{E} (Z^i)^2
$$

$$
= 1 - \mathbb{E} \left[ \frac{1}{n} \sum_{i \in [n]} (1 + Z^i); E_{1/2} \right] - \frac{1}{n} \sum_{i \in [n]} \mathbb{E} (Z^i)^2,
$$

where the first inequality follows by the almost sure nonnegativity of $w_q(B_T)$, the second inequality follows by the almost sure nonnegativity of squares, and the equality because $\mathbb{E}[Z^i] = 0$. Hence,

$$
1 - \mathbb{E} \left[ \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \right] \leq \mathbb{E} \left[ \frac{1}{n} \sum_{i} (1 + Z^i); E_{1/2} \right] + \frac{1}{n} \sum_{i} \mathbb{E} (Z^i)^2 \leq \frac{2L}{p} \exp \left( -\frac{1}{2L} p^2 T \right) + (pT)^{-1}
$$

because Equation (EC.1) implies $1 + Z^i \leq p^{-1}$ a.s., Lemma [EC.12] implies $\mathbb{P}(E_{1/2}^c) \leq 2L \exp(-p^2 T/2)$, and Lemma [EC.11] implies $\mathbb{E}(Z^i)^2 \leq 1/pT$.

**Proof of Theorem 1 for $q \in (-\infty, 0)$**. On the event $E_{1/2} = \{|Z^i| \leq 1/2, \forall i\}$, $B_T > 0$, and

$$
\frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \geq 1 + \frac{1}{q} \sum_{i \in [n]} \left( \frac{\mathbb{E}[B_T]^q}{(1 + Z^i)^q} - 1 \right) \geq 1 + \frac{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q} \frac{Z^i + 2^{-q+2}(Z^i)^2/q}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q}
$$

where the first inequality follows by Lemma [EC.15] and the second by Lemma [EC.16]. We then notice

$$
\mathbb{E} \left[ \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])} \right] \geq \mathbb{E} \left[ \frac{w_q(B_T)}{w_q(\mathbb{E}[B_T])}; E_{1/2} \right]
$$

$$
\geq \mathbb{E} \left[ 1 + \frac{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q (1 + Z^i)}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q} \right] \frac{2^{-q+2} \sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q \mathbb{E}(Z^i)^2}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q}
$$

$$
= 1 - \mathbb{E} \left[ \sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q (1 + Z^i); E_{1/2}^c \right] + \frac{2^{-q+2} \sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q \mathbb{E}(Z^i)^2}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q}
$$

$$
\geq 1 - \mathbb{E} \left[ \max(1 + Z^i); E_{1/2}^c \right] + \frac{2^{-q+2} \sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q \mathbb{E}(Z^i)^2}{\sum_{i \in [n]} (\mathbb{E}[B_T]^q)^q},
$$
where the first inequality follows by the almost sure nonnegativity of \( w_q(B_T) \), the second inequality by above, the third by the almost sure nonnegativity of squares, the equality because \( \mathbb{E}[Z^i] = 0 \), and the last inequality because the coefficients add up to one. Hence,

\[
1 - \mathbb{E} \left[ \frac{w_q(B)}{w_q(\mathbb{E}[B_T])} \right] \leq \mathbb{E} \left[ \max_i (1 + Z^i); E_{1/2}^c \right] + \frac{2^{-q^2}}{q} \max_i \mathbb{E}(Z^i)^2 \leq \frac{2L}{p} \exp \left( -\frac{1}{2L^2} T \right) + \frac{2^{-q^2}}{q} T^{-1}
\]

because Equation (EC.1) implies \( 1 + Z^i \leq p^{-1} \) a.s., Lemma EC.12 implies \( \mathbb{P}(E_{1/2}^c) \leq 2L \exp(-p^2T/2) \), and Lemma EC.11 implies \( \mathbb{E}(Z^i)^2 \leq 1/pT \).

Finally, we conclude by arguing \( w_q(\mathbb{E}[B_T]) = T w_q(\sum_{\ell} p_{\ell} q_{\ell} \xi_{\ell}) \) because \( w_q \) is positively homogeneous for any \( q \in (-\infty, 1) \) and by invoking Lemma 2. □

**EC.1.4. Proof of Lemma 4**

Consider the arrival distribution \( P \) where \( P\{b=(1,1/2)\} = 1/2 = P\{b=(1/2,1)\} \). The hindsight optimum is \( OPT = T/2 - |N_1 - N_2|/6 \). WLOG suppose \( N_1 \leq N_2 \), then this is attained by the offline policy \( \xi_1 = (1,0) \) and \( \xi_2 = (4N_2 - 2T, 2T - N_2)/3N_2 \). The fluid policy is \( (\xi^F)_1 = (1,0) \) and \( (\xi^F)_2 = (0,1) \), which achieves only \( ALG(F) = T/2 - |N_1 - N_2|/2 \). Hence, Lemma EC.13 implies

\[
\mathcal{R}_T(F) = \mathbb{E}[OPT - ALG(F)] = \frac{2}{3} \mathbb{E} \left| N_1 - \frac{T}{2} \right| = \Theta(\sqrt{T}).
\]

**EC.1.5. Proof of Theorem 2**

We start with the telescopic decomposition of the regret:

\[
\mathcal{R}_T(BIRT) = \mathbb{E} [OPT - ALG(BIRT)] = \sum_{k=1}^K \mathbb{E} \left[ OPT^{k-1} - OPT^k \right] + \mathbb{E} \left[ OPT^K - ALG(BIRT) \right], \quad (EC.2)
\]

where we recall \( OPT = OPT^0 \), and \( OPT^k \) is the hindsight optimal egalitarian welfare starting with welfare \( B_{t^*_k} \) at time \( t^*_k \). First notice the last term is subject to an almost sure uniform bound

\[
OPT^K - ALG(BIRT) \leq T - t^*_K = 2, \quad (EC.3)
\]

implied by Lemma EC.1 as if the last re-solving occurs at \( T \), because \( OPT^K \) and \( BIRT \) agree on all actions by \( t^*_K = T - 2 \). Now we proceed with the first terms. For \( k = 1, \ldots, K \), the policies \( OPT^{k-1} \)
and \( \text{OPT}^k \) agree on all allocations by \( t_{k-1}^* \), so Proposition 2 implies (as if time is relabeled to starting at \( t_{k-1}^* \))
\[
E \left[ \text{OPT}^k - \text{OPT}^{k+1} | B_{t_{k-1}^*} \right] \leq 3L(T - t_{k-1}^*) \exp \left( -C \frac{(T - t_{k-1}^*)^4}{(T - t_{k-1}^*)^3} \right).
\]

Notice the constant \( C \) stays the same for all \( k \), as it is determined only by the distribution \( P \) and is independent of time \( T \) and initial welfare \( B_{t_{k-1}^*} \). To simplify the right hand side, we lower bound \( T - t_{k-1}^* \) and upper bound \( T - t_{k-1}^* \) as follows.

First, \( T - t_{k-1}^* = \lfloor \exp(\eta K - k) \rfloor \geq \lfloor e \rfloor = 2 \) implies
\[
\frac{T - t_{k-1}^*}{\exp(\eta K - k)} = \frac{\lfloor \exp(\eta K - k) \rfloor}{\exp(\eta K - k)} \geq \frac{\lfloor \exp(\eta K - k) \rfloor}{\lfloor \exp(\eta K - k) \rfloor + 1} \geq \frac{2}{3}.
\]

Then notice \( T - t_{k-1}^* = \lfloor \exp(\eta K - k) \rfloor \leq \exp(\eta K) \) for \( k = 2, \ldots, K \). For the corner case \( k = 1 \), \( T - t_{0}^* = T \leq \exp(\eta K) \) by definition of \( K = \lfloor \log \log T / \log \eta \rfloor \). Hence, denoting \( C' := (2/3)^4C \), we have for any \( k = 1, \ldots, K \),
\[
E \left[ \text{OPT}^k - \text{OPT}^{k+1} \right] \leq 3Le^{\eta K - k + 1} \exp \left( -C'e^{(4 - 3\eta)\eta K - k} \right).
\]

Then we know
\[
\mathcal{R}_T(\text{BIRT}) \leq 2 + 3L \sum_{k=1}^{K} e^{\eta K - k + 1} \exp \left( -C'e^{(4 - 3\eta)\eta K - k} \right)
\]
\[
= 2 + 3L \sum_{k=1}^{K} e^{\eta k} \exp \left( -C'e^{(4 - 3\eta)\eta k - 1} \right)
\]
\[
\leq 2 + 3L \sum_{k=1}^{K} e^{\eta k} \exp \left( -C'e^{(4\eta - 1)\eta k - 3} \right),
\]

where the first inequality follows from the telescopic decomposition (EC.2) and the almost sure uniform bound (EC.3), the equality by re-ordering the terms, and the last inequality by non-negativity.

Finally, we invoke Lemma EC.17, a result on the sum of the infinite series, to conclude that
\[
\mathcal{R}_T(\text{BIRT}) = E[\text{OPT} - \text{ALG}(\text{BIRT})] \leq 2 + 3LC'^{-\eta / (4 - 3\eta)}h(\eta, 4\eta^{-1} - 3) = O(1).
\]
EC.1.6. Proof of Proposition 2

We first give the following uniform bound on the difference between $\text{OPT}$ and $\text{OPT}^1$, before proceeding to bound the probability that the difference is nonzero. The proof is deferred to Appendix EC.1.7.

**Lemma EC.1.** Suppose all arrival utilities are bounded $b_t \in [0,1]$ for all $t$, and $\text{OPT}^1$ is the hindsight optimal egalitarian welfare starting with utilities $B_{t_1}$ at time $t_1 \in [T]$, with $B_{t_1} \geq B_0$. Then its egalitarian welfare loss against the hindsight optimum starting with utilities $B_0$ at time 0 is at most

$$\text{OPT} - \text{OPT}^1 \leq t_1.$$

Now it remains to bound the decoupling probability $P\{\text{OPT} \neq \text{OPT}^1\}$, because

$$E[\text{OPT} - \text{OPT}^1|B_0] = E[(\text{OPT} - \text{OPT}^1)1\{\text{OPT} \neq \text{OPT}^1\}|B_0] \leq t_1P\{\text{OPT} \neq \text{OPT}^1|B_0\}.$$

We next provide a general sufficient condition on the coupling event $\{\text{OPT} = \text{OPT}^1\}$. For ease of notation, we denote the number of type-$\ell$ arrivals by $N^x_\ell := \sum_{t=1}^{t_1} 1\{b_t = \beta_\ell\}$, and that of those after $t_1$ by $N^y_\ell := \sum_{t=t_1+1}^{T} 1\{b_t = \beta_\ell\}$. We denote using tilted notation the deviation from their respective means (e.g., $\tilde{N}_\ell := N_\ell - E[N_\ell]$). The proof is deferred to Appendix EC.1.8.

**Lemma EC.2.** Suppose the online player acts according to a static policy $\xi^0 \in \Delta^L_n$ for $t = 1,2,\ldots,t_1$. Then $\text{OPT} = \text{OPT}^1$ if there is a hindsight optimal static solution $\xi^* \in \Delta^L_n$ such that

$$N_\ell(\xi^*) \geq N^y_\ell(\xi^0), \quad \forall \ell.$$  \hspace{1cm} (EC.4)

Now that we have a general sufficient condition for $\text{OPT} = \text{OPT}^1$, the remaining parts of the proof can be outlined as the following three steps: 1) invoking sensitivity analysis results for linear programming to show the existence of a hindsight optimal policy that is close to the fluid policy, 2) providing sufficient conditions on the arrivals under which the problem $\text{OPT}$ and $\text{OPT}^1$ are equivalent, and 3) proving that these sufficient conditions hold with high probability using Hoeffding's inequality.
We start by showing that there exists a hindsight optimal policy that is close to the fluid policy, by invoking the LP sensitivity analysis literature against right-hand side constants and exploiting the similarity between the offline problem and the fluid problem. Since the BIRT policy is constructed based on the fluid policy, this guarantees the BIRT policy is close to the hindsight optimal policy, i.e., stochastic fluctuations are small compared to the leeway that the prophet OPT has after time \( t_1 \) to make up for previous suboptimal allocations. Hence, we will be able to show that OPT\(^1\) can recover the exact same utilities for all agents as OPT does with high probability.

We first give a lemma on the proximity between fluid policies and hindsight optimal policies. The proof is deferred to Appendix EC.1.9.

**Lemma EC.3.** For any fluid policy \( \xi^F \), there exists an hindsight optimal policy \( \xi^* \) such that

\[
\max_{i, \ell} \left| N_i(\xi^*)_i^\ell - T p_i(\xi^F)_i^\ell \right| \leq (nL + 1)S \left\| \tilde{\mathbf{N}} \right\|_{\infty},
\]

where \( S \geq 1 \) is a constant determined solely by \( (\beta_i)_{i \in [L]} \), the support of the arrival distribution \( P \).

Now we detail, in Lemma EC.4, sufficient conditions specific to thresholded fluid policies under which Equation (EC.4) is satisfied, so that Lemma EC.2 can be used to argue OPT = OPT\(^1\). The proof can be found in Appendix EC.1.10.

**Lemma EC.4.** Suppose \( \xi^F \) is a fluid policy, and the online player acts \( \xi^0 \) constructed by applying the thresholding rule with \( \gamma \in [0, (T - t_1)/2n^2T] \), for \( t = t_1 + 1, \ldots, T \). Then OPT = OPT\(^1\) if

\[
\tilde{N}_\ell^{\leq} \leq \frac{(T - t_1)p_\ell}{2} \quad \text{for all } \ell, \quad \text{and} \quad \left\| \tilde{\mathbf{N}} \right\|_{\infty} \leq \frac{(T - t_1)p}{2} \frac{\gamma}{S(nL + 1)}, \tag{EC.5}
\]

where we define \( p := \min_\ell p_\ell \).

Using the union bound, Lemma EC.4 implies

\[
P \left\{ \text{OPT} \neq \text{OPT}^1 | B_0 \right\} \leq \sum_{\ell \in [L]} P \left\{ \tilde{N}_\ell^{\leq} > \frac{(T - t_1)p_\ell}{2} \right\} + \sum_{\ell \in [L]} P \left\{ \left| \tilde{N}_\ell \right| > \frac{(T - t_1)p}{2} \frac{\gamma}{S(nL + 1)} \right\}.
\]

We then proceed to invoke Hoeffding’s inequality to bound the probabilities. With \( N_\ell^{\leq} \sim \text{Bin}(t_1, p_\ell) \) and \( N_\ell \sim \text{Bin}(T, p_\ell) \), we have for all \( \ell \in [L] \),

\[
P \left\{ \tilde{N}_\ell^{\leq} > \frac{(T - t_1)p_\ell}{2} \right\} \leq \exp \left( -\frac{(T - t_1)^2}{2t_1} \frac{p_\ell^2}{p^2} \right), \quad \text{and}
\]
\[
P\left\{ |\tilde{N}_t| > \frac{(T-t_1)p}{2} \frac{\gamma}{S(nL+1)} \right\} \leq 2 \exp \left( -\frac{(T-t_1)^2}{2T} \frac{\gamma^2}{S^2(nL+1)^2p^2} \right).
\]

Hence,
\[
P\{ \text{OPT} \neq \text{OPT}^1 | B_0 \} \leq L \left[ \exp \left( -\frac{(T-t_1)^2}{2T} \frac{p^2}{t_1} \right) + 2 \exp \left( -\frac{p^2}{8n^4S^2(nL+1)^2} \frac{(T-t_1)^4}{T^3} \right) \right].
\]

In particular, we can take the threshold \( \gamma = (T-t_1)/2n^2T \) for the tightest bound, so that
\[
P\{ \text{OPT} \neq \text{OPT}^1 | B_0 \} \leq L \left[ \exp \left( -\frac{p^2}{2} \frac{(T-t_1)^2}{t_1} \right) + 2 \exp \left( -\frac{p^2}{8n^4S^2(nL+1)^2} \frac{(T-t_1)^4}{T^3} \right) \right].
\]

Since \( (T-t_1)^2/t_1 \geq (T-t_1)^4/T^3 \) and \( S \geq 1 \), we have
\[
P\{ \text{OPT} \neq \text{OPT}^1 | B_0 \} \leq 3L \exp \left( -\frac{p^2}{8n^4S^2(nL+1)^2} \frac{(T-t_1)^4}{T^3} \right),
\]
and the proof is concluded upon invoking Lemma \[EC.1\].

**EC.1.7. Proof of Lemma [EC.1]**

Suppose a hindsight optimal policy allocates \( x^* \in \Delta_n^T \). Then
\[
\text{OPT} - \text{OPT}^1 = \min_{i \leq n} \left( B_0^i + \sum_{t=1}^T b_t^i(x^*)^i_t \right) - \max_{\{x_t\}_T} \left( B_t^i + \sum_{t=t_1+1}^T b_t^i x_t^i \right)
\leq \min_{i \leq n} \left( B_0^i + \sum_{t=1}^T b_t^i(x^*)^i_t \right) - \min_{i \leq n} \left( B_t^i + \sum_{t=t_1+1}^T b_t^i(x^*)^i_t \right)
= \max_{i \in [n]} \left[ \min_{1 \leq j \leq n} \left( B_0^j + \sum_{t=1}^T b_t^j(x^*)^j_t \right) - \left( B_t^j + \sum_{t=t_1+1}^T b_t^j(x^*)^j_t \right) \right]
\leq \max_{i \in [n]} \left( B_0^i - B_t^i + \sum_{t=1}^{t_1} b_t^i(x^*)^i_t \right)
\leq \max_{i \in [n]} \sum_{t=1}^{t_1} b_t^i(x^*)^i_t
\leq t_1,
\]
where the first equality follows by definition of \( x^* \), the first inequality because \( x^* \in \Delta_n^T \), the second equality because \( -\min = \max \), the second inequality by taking the minimizing agent \( i \) in the latter term, the third inequality because \( B_0 \leq B_{t_1} \), and the last inequality because \( b_t^i \leq 1 \) and \( (x^*)^i_t \leq 1 \).
EC.1.8. Proof of Lemma EC.2

Clearly, \( \text{OPT} = \text{OPT}^1 \) if for any arrival type, any agent can get as much utility from \( \text{OPT}^1 \)'s policy as from \( \text{OPT} \)'s policy, i.e. there exists \( \xi \in \Delta_n^L \) such that

\[
N_i(\xi^*)_\ell = N^\xi_\ell(\xi^0)_\ell + N^\xi_\ell \xi^*_\ell.
\]

Fix arrival type \( \ell \). If \( N^\xi_\ell = 0 \), then we claim Equation (EC.4) holds with equality. Otherwise,

\[
N_\ell = \sum_i N_i(\xi^*)_\ell > \sum_i N^\xi_i(\xi^0)_\ell = N^\xi_\ell = N_\ell,
\]

which is absurd. Now we have shown \( N_i(\xi^*)_\ell = N^\xi_\ell(\xi^0)_\ell \) for all \( i \), i.e., all agents have accumulated the exact welfare needed from type-\( \ell \) arrivals by time \( t_1 \).

An arbitrary allocation for type \( \ell \) would satisfy the condition, since there will be no such arrivals.

If \( N^\xi_\ell > 0 \), the proof is concluded by simply taking the static policy \( \xi \in \Delta_n^L \) given by

\[
\xi_\ell = \frac{N_i(\xi^*)_\ell - N^\xi_\ell(\xi^0)_\ell}{N^\xi_\ell}, \quad \forall \ell \in [L].
\]

EC.1.9. Proof of Lemma EC.3

We restate the following result on sensitivity of optimal solutions against RHS in linear programs.

**Lemma EC.5** (Schrijver (1998), Theorem 10.5). Let \( A \) be an \( p \times m \)-matrix, and let \( S \) be such that for each nonsingular submatrix \( M \) of \( A \) all entries of \( M^{-1} \) are at most \( S \) in absolute value. Let \( c \) be a row \( n \)-vector, and let \( b' \) and \( b'' \) be column \( m \)-vectors such that \( \max\{cx | Ax \leq b'\} \) and \( \max\{cx | Ax \leq b''\} \) are finite. Then for each optimum solution \( x' \) of the first maximum there exists an optimum solution \( x'' \) of the second maximum with

\[
\|x' - x''\|_\infty \leq mS \|b' - b''\|_\infty.
\]

To apply the sensitivity analysis result, we rewrite in the linear program form the offline problem as (with optimal solution \( (y^*)_\ell = N_\ell(\xi^*)_\ell \))

\[
\text{OPT} = \max \quad U
\]

s.t. \[
U \leq B^i_0 + \sum_\ell \beta^i_\ell y^i_\ell, \quad \forall i,
\]

\[
y^i_\ell \geq 0, \quad \forall i, \ell,
\]

\[
\sum_i y^i_\ell = N_\ell, \quad \forall \ell,
\]
and the initial fluid problem can be rewritten as (with optimal solution \((z^*)^i_\ell := Tp_\ell(\xi^F)^i_\ell\))

\[
\text{FLU} = \max \quad U \\
\text{s.t.} \quad U \leq B_0^i + \sum_\ell \beta^i_\ell z^i_\ell, \quad \forall i, \\
\quad z^i_\ell \geq 0, \quad \forall i, \ell, \\
\quad \sum_i z^i_\ell = Tp_\ell, \quad \forall \ell.
\]

Lemma [EC.5] implies for any optimal solution \(z\) to the fluid problem, there exists an optimal solution \(y\) to the offline problem such that

\[
\max\{|\text{FLU} - \text{OPT}|, \|y - z\|_\infty\} \leq (nL + 1)S\left\|\tilde{N}\right\|_\infty,
\]

where \(S\) is larger than any absolute value of entries in the inverse of any nonsingular submatrix of

\[
A = \begin{bmatrix}
1_n & \text{diag}(-\beta_1) \cdots \text{diag}(-\beta_L) \\
0_{n\times L} & -I_{n\times L} \\
0_L & 1_n^\top \cdots 0 \\
& \vdots \ddots \vdots \\
0 & \cdots 1_n^\top \\
0_L & -1_n^\top \cdots 0 \\
& \vdots \ddots \vdots \\
0 & \cdots -1_n^\top
\end{bmatrix}.
\]

Here the decision variable \(x\) in the statement of the lemma is \(x = (U, y_1, \ldots, y_L)\) or \(x = (U, z_1, \ldots, z_L)\) and the right-hand side vector of the constraints is \((B_0, 0_{n\times L}, \tilde{N}, -\tilde{N})\) or \((B_0, 0_{n\times L}, T\tilde{p}, -T\tilde{p})\), respectively. In particular, we also observe \(S \geq 1\). The proof is concluded as a corollary.

**EC.1.10. Proof of Lemma [EC.4]**

First, Lemma [EC.3] implies the existence of a hindsight optimal static policy \(\xi^* \in \Delta^L_n\) such that

\[
\max_{i, \ell} |N_\ell(\xi^*)_i - Tp_\ell(\xi^F)_i| \leq (nL + 1)S\left\|\tilde{N}\right\|_\infty.
\]
so for all $i$ and $\ell$,

$$N_i(\xi^*)_\ell - N^\xi_f(\xi^f)_\ell \geq \left[ Tp_{\ell}(\xi^f)_\ell - (nL + 1)S \left\| \tilde{N} \right\|_\infty \right] - \left[ t_1p_{\ell} + N^\xi_f \right] (\xi^f)_\ell$$

$$= \left[ (T - t_1)p_{\ell} - \tilde{N}^\xi_f \right] (\xi^f)_\ell - (nL + 1)S \left\| \tilde{N} \right\|_\infty.$$

Lemma EC.2 implies that it suffices to show $N_i(\xi^*)_\ell \geq N^\xi_f(\xi^0)_\ell$ for all $i$ and $\ell$. Now we check condition EC.4 element-wise. Recall from the thresholding rule that $(\xi^0)_\ell \in \{0\} \cup [\gamma, 1]$ for all $i$ and $\ell$.

- If $(\xi^0)_\ell = 0$, then $N_i(\xi^*)_\ell \geq 0 = N^\xi_f(\xi^0)_\ell$.
- If $(\xi^0)_\ell \geq \gamma$ and $i \neq j$, then $(\xi^f)_\ell = (\xi^0)_\ell \geq \gamma$. We have

$$N_i(\xi^*)_\ell - N^\xi_f(\xi^0)_\ell = N_i(\xi^*)_\ell - N^\xi_f(\xi^f)_\ell$$

$$\geq \left[ (T - t_1)p_{\ell} - \tilde{N}^\xi_f \right] (\xi^f)_\ell - (nL + 1)S \left\| \tilde{N} \right\|_\infty$$

$$\geq \frac{(T - t_1)p_{\ell} (\xi^f)_\ell}{2} - \frac{(T - t_1)p_{\ell}}{2} \gamma$$

$$\geq 0,$$

where the equality follows because $(\xi^f)_\ell = (\xi^0)_\ell$, the first inequality by Equation (EC.6), the second by assumption EC.5, and the last because $(\xi^f)_\ell = (\xi^0)_\ell \geq \gamma$.

- Agent $j \in \arg\max_i (\xi^f)_\ell$ is the only one receiving positive adjustment. We have

$$N_i(\xi^*)_\ell - N^\xi_f(\xi^0)_\ell = N_i(\xi^*)_\ell - N^\xi_f(\xi^f)_\ell - N^\xi_f(\xi^0 - \xi^f)_\ell$$

$$\geq N_i(\xi^*)_\ell - N^\xi_f(\xi^f)_\ell - n\gamma N^\xi_f$$

$$\geq \left[ Tp_{\ell}(\xi^f)_\ell - (nL + 1)S \left\| \tilde{N} \right\|_\infty \right] - N^\xi_f(\xi^f)_\ell - n\gamma N^\xi_f$$

$$= \left[ (T - t_1)p_{\ell} - \tilde{N}^\xi_f \right] \left[ (\xi^f)_\ell + n\gamma \right] - (nL + 1)S \left\| \tilde{N} \right\|_\infty - n\gamma Tp_{\ell}$$

$$\geq \frac{(T - t_1)p_{\ell}}{2} \left[ (\xi^f)_\ell + n\gamma \right] - \frac{(T - t_1)p_{\ell}}{2} \gamma - n\gamma Tp_{\ell}$$

$$\geq nTp_{\ell} \left[ \frac{T - t_1}{2n^2T} - \gamma \right] + \frac{(T - t_1)p_{\ell}}{2} (n - 1)\gamma$$

$$\geq 0.$$

where the first equality follows by rearranging terms, the first inequality because the adjustment is at most $(\xi^0 - \xi^f)_\ell = -\sum_{i \neq j} (\xi^0 - \xi^f)_i < n\gamma$, the second inequality by Equation (EC.6), the second
equality because \( N_t^\leq = t_1 p_\leq + \tilde{N}_x^\leq \), the third inequality by assumption (EC.5), the fourth inequality because \((\xi^ f)^i \geq \max_i(\xi^{ f})^i \geq 1/n\), and the last inequality by the threshold rule \( 0 \leq \gamma_0 \leq (T-t_1)/2n^2T \).

Therefore, we conclude \( N_t(\xi^0)^i \geq N_t^\leq (\xi^{ f})^i \) for all \( i \) and \( \ell \), and so \( \text{OPT} = \text{OPT}^1 \) by Lemma EC.2.

EC.2. Additional results

**Lemma EC.6 (Futility of randomization).** Consider the online convex optimization problem

\[
\text{maximize } w \left( \sum_{t=1}^T b_t x_t \right)
\]

subject to \( x_t \in \mathcal{X}_t \),

\[
x_t \perp \sigma(b_{t+1}, \ldots, b_T)
\]

where \( w : \mathbb{R}^m \to \mathbb{R} \) is a concave function, \( b_t \in \mathbb{R}^{m \times n} \) are fixed arrivals, and \( \mathcal{X}_t \subseteq \mathbb{R}^n \) is a convex set for every \( t \). Then in expectation, any randomized online policy is weakly dominated by some deterministic online policy.

**Proof of Lemma EC.6** Fix any randomized online policy \( \pi \), and suppose it outputs random actions \((x_t(\pi) : t \in [T])\). Jensen’s inequality implies that

\[
\mathbb{E}_\pi \left[ w \left( \sum_{t=1}^T b_t x_t(\pi) \right) \right] \leq w \left( \sum_{t=1}^T b_t \mathbb{E}_\pi [x_t(\pi)] \right).
\]

Notice the mean policy \( \bar{\pi} \) given by actions \( x_t(\bar{\pi}) = \mathbb{E}_\pi [x_t(\pi)] \) is feasible, because \( x_t(\bar{\pi}) \in \mathcal{X}_t \) by the convexity of the action domains \( \mathcal{X}_t \), so it is an online policy, and the RHS is simply its objective.

\( \square \)

**Lemma EC.7 (Static policies dominate).** \( P \)-almost surely, any offline policy is weakly dominated by a static policy.

**Proof of Lemma EC.7** Based on any offline policy \( \mathbf{x} \), the proof is immediate by defining the static policy given by

\[
\xi_\ell := \frac{1}{N_\ell} \sum_{t : b_t = \beta_\ell} x_t \quad \text{for all } \ell \in [L] \text{ where } N_\ell > 0,
\]

because this static policy satisfies \( \xi_\ell \in \Delta_n \) and gives agents the same utilities as the offline policy does. \( \square \)
Lemma EC.8. \( P \)-almost surely, there is a hindsight optimal policy under which the agents have equal utilities.

**Proof of Lemma EC.8** Fix an arrival sequence \( b \equiv (b_t : t \in [T]) \). We first deal with the corner case \( \text{OPT} = 0 \). This means that a naïve even allocation \( (x_t^i = 1/n \text{ for all } t \text{ and } i) \) also attains zero welfare, so there is some agent \( i^* \) where \( b_t^{i^*} = 0 \) for all \( t \). Hence, allocating all resources to agent \( i^* \) results in all agents receiving zero utility.

In the sequel suppose the hindsight optimum \( \text{OPT} > 0 \) is achieved by some optimal offline allocation sequence \( x \) that result in unequal utilities across agents. Then we can denote the set of agents with excess utilities as \( I := \{i : \sum_t b_t^i x_t^i > \text{OPT}\} \neq \emptyset \). Notice by definition, \( \sum_t b_t^i x_t^i \geq \text{OPT} \) for all \( i \in [n] \), so \( \sum_t b_t^i x_t^i = \text{OPT} \) for \( i \in [n] \setminus I \). We aim to construct an alternative allocation sequence that equalize all agents’ utilities while attaining the hindsight optimum \( \text{OPT} \).

We first consider the heuristic allocation sequence \( \hat{x} \) given by

\[
\hat{x}_t^i := \frac{\text{OPT}}{\sum_t b_t^i x_t^i} x_t^i + \frac{1}{n} \sum_{j \in [n]} \left(1 - \frac{\text{OPT}}{\sum_t b_t^j x_t^j}\right) x_t^j, \quad i = 1, \ldots, n.
\]

Clearly \( \hat{x}_t \in \Delta_n \), because \( \hat{x}_t \geq 0 \) and \( \sum_i \hat{x}_t^i = 1 \). Then under \( \hat{x} \), agent \( i \) would have a cumulative utility of

\[
\sum_{t=1}^T b_t^i \hat{x}_t^i = \text{OPT} + \frac{1}{n} \sum_{j \in [n]} \left(1 - \frac{\text{OPT}}{\sum_t b_t^j x_t^j}\right) \sum_{t=1}^T b_t^j x_t^j.
\]

By definition, \( \hat{x} \) cannot perform better than the optimal offline policy, so \( \min_{1 \leq i \leq n} \sum_{t=1}^T b_t^i \hat{x}_t^i \leq \text{OPT} \).

Since \( \sum_t b_t^i x_t^i \geq \text{OPT} \) for all \( i \) with strict inequality holding for \( j \in I \), this implies there exists some agent \( i^* \in [n] \) such that

\[
b_t^{i^*} x_t^{i^*} = 0, \quad \forall j \in I, t \in [T]. \tag{EC.7}
\]

This necessary condition \( \tag{EC.7} \) is similar to the well-known complementary slackness condition in duality theory. It fixes an agent \( i^* \) and stipulates that resources with positive utility to \( i^* \) \( (t \text{ when } b_t^{i^*} > 0) \) may not be allocated to agents with excess utilities, i.e., \( x_t^j = 0 \) for all \( j \in I \). In particular, we notice that \( i^* \notin I \), since otherwise \( b_t^{i^*} x_t^{i^*} = 0 \) for all \( t \), contradicting \( \sum_t b_t^{i^*} x_t^{i^*} \geq \text{OPT} > 0 \). This means \( i^* \) is a binding agent, i.e., \( \sum_t b_t^{i^*} x_t^{i^*} = \min_{1 \leq i \leq n} \sum_t b_t^i x_t^i = \text{OPT} \).
After we pick out the agent $i^*$, condition (EC.7) implies that any positive allocation $x^i_j > 0$ to agents with excess utilities $j \in I$ can be safely redirected toward agent $i^*$ without increasing its cumulative utility.

More formally, consider the allocation sequence $\tilde{x}$ given by

$$\tilde{x}^i_t := \frac{\OPT}{\sum_t b^i_t} x^i_t, \quad \forall i \neq i^*,$$

$$\tilde{x}^{i^*}_t := x^{i^*}_t + \sum_{j \in [n]} \left(1 - \frac{\OPT}{\sum_t b^j_t x^j_t}\right) x^j_t.$$

Clearly $\tilde{x} \in \Delta_n$, because $\tilde{x} \geq 0$, and $\sum_i x^i_t = 1$. Furthermore, it is in fact a resource-equalizing hindsight optimal solution, because

$$\sum_{t=1}^T b^i_t \tilde{x}^i_t = \sum_{t=1}^T b^i_t \frac{\OPT}{\sum_t b^i_t} x^i_t = \OPT, \quad \forall i \neq i^*,$$

$$\sum_{t=1}^T b^{i^*}_t \tilde{x}^{i^*}_t = \sum_{t=1}^T b^{i^*}_t x^{i^*}_t + \sum_{j \in [n]} \left(1 - \frac{\OPT}{\sum_t b^j_t x^j_t}\right) \sum_{t=1}^T b^j_t x^j_t = \sum_t b^i_t x^{i^*}_t = \OPT,$$

where the last inequality follows because $i^* \notin I$. Hence we have constructed a hindsight optimal allocation that equalize the agents’ cumulative utilities. □

**Lemma EC.9.** There exists some $P \in \mathcal{P}$ under which $\text{FLU} - \mathbb{E} [\OPT] = \Omega(\sqrt{T})$.

**Proof of Lemma EC.9** Take a simple example where $n = 2$ agents start with zero utilities $B_0 = (0, 0)$, and the arrival structure is given by $\beta_1 = (1, 0)$ and $\beta_2 = (0, 1)$ with respective probabilities $p_1 = \frac{1}{2}$ and $p_2 = \frac{1}{2}$. Clearly the optimal value of the fluid problem is $\text{FLU} = \frac{T}{2}$, attained by the fluid policy $\xi_1 = (1, 0)$ and $\xi_2 = (0, 1)$.

To compute $\OPT$, note the trivial bound $\OPT \leq \min(\sum_{t=1}^T b^1_t, \sum_{t=1}^T b^2_t) = \min(N_1, N_2)$; in fact, it can be attained by the naïve policy $x_t = b_t$, so $\OPT = \min(N_1, N_2)$. We then know

$$\text{FLU} - \mathbb{E} [\OPT] = \mathbb{E} \left| N_1 - \frac{T}{2} \right| = \Theta(\sqrt{T}),$$

by invoking Lemma EC.13 in Appendix EC.3 □

**Lemma EC.10.** Under Hölder-mean welfare $w_q$ for $q \in [-\infty, 1]$ and any static policy $\xi$, $w_q(B_T) \leq T/n$. 
Proof of Lemma EC.10. Under static policy $\xi$,\[
\begin{aligned}
    w_q(B_T) &= w_q \left( \sum_{\ell \in [L]} N_\ell \beta_\ell \ast \xi_\ell \right) \\
    &\leq w_q \left( \sum_{\ell \in [L]} N_\ell \xi_\ell \right) \\
    &\leq w_q \left( 1 \sum_{i \in [n]} \sum_{\ell \in [L]} N_\ell \xi_\ell^i \right) \\
    &= \frac{T}{n},
\end{aligned}
\]
where the first equality follows by definition, the first inequality because $\beta_\ell \in [0,1]^n$, the second inequality because $w$ is concave, and the second equality because $\sum_\ell N_\ell = T$, $\xi_\ell \in \Delta_n$ and $w_q(1) = 1$.

**Lemma EC.11.** For any static policy $\xi$ and agent $i$,\[
\begin{aligned}
    \text{Var}(B_T^i) &\leq \frac{\mathbb{E}[B_T^i]^2}{T}.
\end{aligned}
\]

*Proof of Lemma EC.11.* First notice \[
\begin{aligned}
    \text{Var}(B_T^i) &= \sum_\ell \text{Var}(N_\ell) (\beta_\ell^i \xi_\ell^i)^2 + \sum_{\ell \neq \ell'} \text{Cov}(N_\ell, N_{\ell'}) \beta_\ell^i \beta_{\ell'}^i \xi_\ell^i \xi_{\ell'}^i \\
    &= T \sum_\ell p_\ell (1 - p_\ell) (\beta_\ell^i \xi_\ell^i)^2 + T \sum_{\ell \neq \ell'} (-p_\ell p_{\ell'}) \beta_\ell^i \beta_{\ell'}^i \xi_\ell^i \xi_{\ell'}^i \\
    &= T \sum_\ell p_\ell (\beta_\ell^i \xi_\ell^i)^2 - T \left( \sum_\ell p_\ell \beta_\ell^i \xi_\ell^i \right)^2 \\
    &\leq T \sum_\ell p_\ell (\beta_\ell^i \xi_\ell^i)^2,
\end{aligned}
\]
where the first equality follows because $B_T^i = \sum_\ell N_\ell \beta_\ell^i \xi_\ell^i$ by definition, the second equality because $(N_\ell)_{\ell} \sim \text{Multinomial}(T, (p_\ell)_{\ell})$, the third equality by rearranging terms, and the inequality by non-negativity of squares. Next, note \[
\begin{aligned}
    \mathbb{E}[B_T^i(\xi)]^2 &= T^2 \left( \sum_\ell p_\ell \beta_\ell^i \xi_\ell^i \right)^2 \\
    &\geq T^2 \sum_\ell (p_\ell \beta_\ell^i \xi_\ell^i)^2 \\
    &\geq p T^2 \sum_\ell p_\ell (\beta_\ell^i \xi_\ell^i)^2,
\end{aligned}
\]
where the equality follows by definition, the first inequality follows by expanding the squares, and the second inequality because $p_\ell \geq p$ for all $\ell$. The proof is thereby concluded. $\Box$
Lemma EC.12. Given any $T \geq 1$ and static policy $\xi$, define the event $E_{\delta} := \{|B^i_T| \leq \delta E[B^i_T], \forall i\}$ for all $\delta > 0$. Then $\mathbb{P}(E^c_{\delta}) \leq 2L \exp(-2p^2T\delta^2)$, where $p := \min_{\ell} p_{\ell}$.

Proof of Lemma [EC.12]

$$
\mathbb{P}(E^c_{\delta}) = \mathbb{P}\{|B^i_T| > \delta E[B^i_T] \text{ for some } i\}
\leq \mathbb{P}\{|\tilde{N}_\ell| > \delta E[N_\ell] \text{ for some } \ell\}
\leq \sum_{\ell \in [L]} \mathbb{P}\{|\tilde{N}_\ell| > \delta E[N_\ell]\}
\leq \sum_{\ell \in [L]} 2 \exp\left(-\frac{2(\delta E[N_\ell])^2}{T}\right)
= \sum_{\ell \in [L]} 2 \exp\left(-2p^2T\delta^2\right)
\leq 2L \exp(-2p^2T\delta^2),
$$

where the first inequality follows because $|\tilde{N}_\ell| \leq \delta E[N_\ell]$ for all $\ell$ implies $|B^i_T| \leq \delta E[B^i_T]$ for all $i$, the second by the union bound, the third by Hoeffding’s inequality, and the last because $p_{\ell} \geq p$. □

We give a generalization of Theorem 2 below, allowing for dependency of the distribution on time horizon length $T$.

Definition EC.1. Given a finite support $\beta \equiv (\beta_\ell \in [0, 1]^n : \ell \in [L])$ and a positive $p \in (0, \frac{1}{2})$, the class of restricted admissible probability distributions $P(p, \beta)$ is the set of all finite probability distributions $P$ with support $\beta$ such that $P(b = \beta_\ell) \geq p$ for all $\ell \in [L]$.

Theorem EC.1. (Extension of Theorem 2) Fix $\epsilon \in (0, \frac{1}{2})$. Fix any sequence $p(T) \in (0, \frac{1}{2})$ with $p(T) = \Omega(T^{-1/2+\epsilon})$. Then for any $\eta \in (1, (1 - \epsilon/2)^{-1})$,

$$
\sup_{P \in P(p(T), \beta)} R_T(B^\eta RT) = O(1). \quad (EC.8)
$$

Proof of Theorem [EC.1] By assumption there exist $T_0 > 1$ and $\sigma > 0$ such that $p(T) \geq \sigma T^{-1/2+\epsilon}$ for all $T > T_0$. It suffices to assume $T > T_0$. Recall Proposition[2] shows

$$
\mathbb{P}\{OPT \neq OPT^1\} \leq 3L \exp\left(-\frac{p^2}{8n^4S^2(nL + 1)^2} \frac{(T - t_1)^4}{T^3}\right)
\leq 3L \exp\left(-\frac{\sigma^2}{8n^4S^2(nL + 1)^2} \frac{(T - t_1)^4}{T^{4-2\epsilon}}\right).
$$
In particular, for re-solving points defined in the BIRT\(\eta\) policy, we can show, similar to the proof of Theorem 2 (Appendix EC.1.5), that
\[
\mathbb{E} \left[ \text{OPT}^{k-1} - \text{OPT}^k \right] \leq 3Le^{\eta}K^{-k+1} \exp \left( -C''e^{(4-(4-2\varepsilon)\eta)K^{-k}} \right),
\]
for some constant \(C'' > 0\) determined by \(\sigma, T_0,\) and the support of the distribution. Finally, we invoke Lemma EC.17 to conclude, following a similar convergence argument as in the proof of Theorem 2 (Appendix EC.1.5), that for any \(P \in \mathcal{P}(p(T), \beta)\) and any choice of \(\eta \in (1, (1-\varepsilon/2)^{-1}),\)
\[
\mathcal{R}_T(\text{BIRT}) \leq 2 + 3LC''^{-1/\zeta}h(\eta, \zeta) = O(1),
\]
where \(\zeta = 4\eta^{-1} - (4-2\varepsilon) \in (0, 1). \quad \Box

**EC.3. Additional proofs**

**Lemma EC.13.** Suppose \(N_T \sim \text{Bin}(T, 1/2)\) for \(T \geq 1\). Then
\[
\frac{\mathbb{E}[|N_T - T/2|]}{\sqrt{T}} \to \frac{1}{\sqrt{2\pi}} \quad \text{as} \quad T \to \infty.
\]

*Proof of Lemma EC.13* Define the sequence \(M_T := (2N_T - T)/\sqrt{T}\) for \(T \geq 1\). Then it suffices to show that \(\mathbb{E}|M_T| \to \sqrt{2/\pi}\) as \(T \to \infty\).

Observe that \(\mathbb{E}M_T = 0\) and \(\mathbb{E}M_T^2 = 4\text{Var}(N_T)/T = 1\) for all \(T\), so \(\{M_T\}\) is uniformly integrable, and that the Central Limit Theorem implies \(M_T \Rightarrow N(0, 1)\). By Skorokhod’s Representation Theorem, there exists a sequence of random variables \((Z_T : T \geq 1)\) such that \(M_T \overset{d}{=} Z_T\) and \(Z_T \to Z\) almost surely, where \(Z \sim N(0, 1)\). This means the class \(\{Z_T : T \geq 1\}\) is also uniformly integrable, so \(Z_T \to Z\) in \(L^1\), implying in particular \(\mathbb{E}|Z_T| \to \mathbb{E}|Z| = \sqrt{2/\pi}\). Hence, we conclude that \(\mathbb{E}|M_T| = \mathbb{E}|Z_T| \to \sqrt{2/\pi}. \quad \Box

**Lemma EC.14.** For any \(x \geq -1/2\), \(\log(1 + x) \geq x - x^2\).

*Proof of Lemma EC.14* Denote the smooth function \(f \in C^\infty([-1, \infty))\) given by \(f(x) = \log(1 + x) - x + x^2\), then \(f(0) = 0\) and \(f(-1/2) = 3/4 - \log 2 > 0\). The derivative of \(f(x)\) is
\[
f'(x) = (1 + x)^{-1} - 1 + 2x,
\]
so \( f'(0) = 0 \) and \( f'(-1/2) = 0 \). Next, we argue that 0 is the only local minimum of \( f \). The second-order derivative of \( f \) is

\[
  f''(x) = -(1 + x)^{-2} + 2.
\]

Note it is monotonically increasing in \((-1, \infty)\); in particular, \( f'' < 0 \) on \((-1, -1+1/\sqrt{2})\) and \( f'' > 0 \) on \((-1+1/\sqrt{2}, +\infty)\). Now \( f'(x) = f'(0) + \int_0^x f''(z) \, dz \) for any \( x \in (-1/2, \infty) \), so for any \( x \neq 0 \) where \( f'(x) = 0 \), \( f''(x) < 0 \). This means it cannot be a local minimum. Hence, 0 is the only local minimum of \( f \).

Finally, notice \( f'(0) = 0 \) and \( f''(x) > 0 \) for any \( x \geq 0 \) implies that \( f(x) > f(0) = 0 \) for any \( x \geq 0 \). Since we know \( f(0) = 0 \), we conclude that 0 is a global minimum of \( f \) and that \( f(x) \geq 0 \) for all \( x \geq -1/2 \). □

**Lemma EC.15.** For \( x \geq -1 \),

\[
  (1 + x)^q \geq \begin{cases} 
  1 + qx - (1 - q)x^2, & q \in (0, 1), \\
  1 + qx, & o/w.
\end{cases}
\]

**Proof of Lemma EC.15** Suppose \( q \in (0, 1) \) and denote the smooth function \( f \in C^\infty([-1, \infty)) \) given by \( f(x) = (1 + x)^q - 1 - qx + (1 - q)x^2 \), then \( f(-1) = 0 \) and \( f(0) = 0 \). The derivative of \( f(x) \) is

\[
  f'(x) = q(1 + x)^{q-1} - q + 2(1 - q)x,
\]

so \( f'(0) = 0 \) and \( f'(-1^+) = +\infty \). Next, we argue that 0 is the only local minimum of \( f \). The second-order derivative of \( f \) is

\[
  f''(x) = q(q - 1)(1 + x)^{q-2} + 2(1 - q).
\]

Note it is monotonically increasing in \((-1, \infty)\), \( f''(-1^+) = -\infty \), and \( f''(0) = q^2 - 2q + 2 > 0 \). Now \( f'(x) = f'(0) + \int_0^x f''(z) \, dz \) for any \( x \in [-1, \infty) \), so for any \( x \neq 0 \) where \( f'(x) = 0 \), \( f''(x) < 0 \). This means it cannot be a local minimum. Hence, 0 is the only local minimum of \( f \).
Finally, notice $f'(0) = 0$ and $f''(x) > 0$ for any $x \geq 0$ implies that $f(x) > f(0) = 0$ for any $x \geq 0$. Since we know $f(0) = 0$, we conclude that 0 is a global minimum of $f$ and that $f(x) \geq 0$ for all $x \geq -1$.

Now suppose $q \in (-\infty, 0] \cup [1, \infty)$ and consider the smooth function $g \in C^\infty([-1, \infty))$ given by $g(x) = (1 + x)^q - (1 + qx)$. Then $f(0) = 0$. The first-order derivative of $g$ is $g'(x) = q(1 + x)^{q-1} - q$, so $g'(0) = 0$. The second-order derivative of $g$ is $g''(x) = q(q-1)(1 + x)^{q-2} \geq 0$ for all $x \in (-1, \infty)$ because $q \not\in (0, 1)$. This means $g'$ is nonpositive on $(-1, 0)$ and nonnegative on $(0, \infty)$, so $g$ attains its minimum at 0. □

**Lemma EC.16.** For $x \geq -1/2$ and $q < 0$,

$$(1 + x)^q \leq 1 + qx + 2^{-q+2}x^2.$$ 

**Proof of Lemma EC.16** Denote the smooth function $f \in C^\infty([-1, \infty))$ given by $f(x) = (1 + x)^q - 1 - qx - 2^{-q+2}x^2$, then $f(-1/2) = -1 + q/2 < 0$ and $f(0) = 0$. The derivative of $f(x)$ is

$$f'(x) = q(1 + x)^{q-1} - q - 2^{-q+3}x,$$

so $f'(0) = 0$. Next, we argue that 0 is the only local maximum of $f$. The second-order derivative of $f$ is

$$f''(x) = q(q-1)(1 + x)^{q-2} - 2^{-q+3}.$$ 

Note it is monotonically decreasing in $(-1, \infty)$ and $f''(0) = q(q-1) - 2^{-q+3} < 0$. Now $f'(x) = f'(0) + \int_0^x f''(z) \, dz$ for any $x \in [-1, \infty)$, so for any $x \neq 0$ where $f'(x) = 0$, $f''(x) > 0$. This means it cannot be a local maximum. Hence, 0 is the only local maximum of $f$.

Finally, notice $f'(0) = 0$ and $f''(x) < 0$ for any $x \geq 0$ implies that $f(x) < f(0) = 0$ for any $x \geq 0$. Since we know $f(0) = 0$, we conclude that 0 is a global minimum of $f$ and that $f(x) \geq 0$ for all $x \geq -1$. □

**Lemma EC.17.** Given $\eta > 1$ and $\zeta \in (0, 1)$, for any $C > 0$,

$$\sum_{k=1}^\infty e^{\eta^k} \exp \left(-C e^{\eta^k \zeta}\right) \leq C^{-1/\zeta} h(\eta, \zeta),$$
where

\[ h(\eta, \zeta) := \frac{1}{\zeta \log \eta} \Gamma \left( \frac{1}{\zeta} \right) + 2\zeta^{-1/\zeta}. \]

**Proof of Lemma EC.17** For ease of notation, we define a function \( a : \mathbb{R} \to \mathbb{R} \) given by \( a(w) = \eta^w - C \eta^w \zeta \) so that the LHS of the inequality we are to show is \( \sum_{k=1}^{\infty} e^{a(k)} \). Clearly \( a(\cdot) \) is continuously differentiable in \((0, \infty)\), and its first-order derivative is

\[ a'(w) = \left( 1 - C \zeta \eta^w \zeta \right) \eta^w \log \eta. \]

If \( C \zeta \geq 1 \), then \( a'(w) < 0 \) for all \( w > 0 \), so that \( a(\cdot) \) is monotonically decreasing, and \( \sum_{k=1}^{\infty} e^{a(k)} \leq \int_0^{\infty} e^{a(w)} \, dw \). Otherwise, \( a'(w) > 0 \) and \( a(w) \) is increasing for \( w \in (0, w^*) \), and \( a'(w) < 0 \) and \( a(w) \) is decreasing for \( w \in (w^*, \infty) \), where

\[ w^* := \log \left( \frac{-1}{\log(C \zeta)} \right) > 0. \]

This means \( a(\cdot) \) is maximized at \( w^* \), where it takes value

\[ a(w^*) = \frac{-\log(C \zeta) - 1}{\zeta}. \]

Denoting \( k^* := \lfloor w^* \rfloor \geq 0 \), we have

\[
\sum_{k=1}^{\infty} e^{a(k)} \leq \sum_{k=1}^{k^*-1} e^{a(k)} + e^{a(k^*)} + \sum_{k=k^*+2}^{\infty} e^{a(k)} \\
\leq \int_0^{k^*} e^{a(w)} \, dw + 2e^{a(w^*)} + \int_{k^*+1}^{\infty} e^{a(w)} \, dw \\
\leq \int_0^{\infty} e^{a(w)} \, dw + 2(C \zeta)^{-1/\zeta},
\]

where in the first inequality we assume \( \sum_{k=p}^{q} e^{a(k)} = 0 \) for any \( p > q \), the second follows because \( a(w) \) increases in \((0, k^*)\) and decreases in \((k^* + 1, \infty)\), and the last by definition of \( w^* \).
Now it remains to bound the integral \( \int_0^\infty e^{a(w)} \, dw \). In fact, we conclude using change of variables that

\[
\int_0^\infty e^{a(w)} \, dw = \int_0^\infty \exp \left( \eta w - C e^{\eta w / \zeta} \right) \, dw \\
= \int_1^\infty \exp \left( u - C e^{u / \zeta} \right) \frac{1}{u \log \eta} \, du \\
\leq \frac{1}{\log \eta} \int_1^\infty \exp \left( u - C e^{u / \zeta} \right) \, du \\
\leq \frac{1}{\log \eta} \int_0^\infty \exp \left( u - C e^{u / \zeta} \right) \, du \\
= \frac{C^{-1/\zeta}}{\log \eta} \int_0^\infty \exp \left( -v^\zeta \right) \, dv \\
= C^{-1/\zeta} \frac{\Gamma(1/\zeta)}{\zeta \log \eta},
\]

where the first equality is by definition of \( a(\cdot) \), the second equality by change of variables \( u = \eta w \), the first inequality because \( u \geq 1 \), the second inequality by non-negativity, the third equality by change of variables \( v = C^{-1/\zeta} e^u \), and the last equality by definition of the Gamma function \( \Gamma(\cdot) \). □
EC.4. Additional policies

Algorithm 3: Backward Infrequent Re-solving (BIR$_\eta$)

**Input:** time horizon length $T \in \mathbb{N}$, initial utilities $B_0 \in \mathbb{R}^n$, hyper-parameter $\eta \in (1, 4/3)$.

**Initialize:** set $K \leftarrow \lceil \log \log T / \log \eta \rceil$;

set $t^*_0 \leftarrow 0$, $t^*_k \leftarrow T - \lceil \exp(\eta^{K-k}) \rceil$ for $k = 1, \ldots, K$, and $t^*_{K+1} \leftarrow T$;

for $k = 0, 1, \ldots, K$ do

solve the fluid problem with updated utilities:

$$\xi^F \in \arg \max \left\{ w \left( B_{t^*_k} + (T - t^*_k) \sum_{\ell \in [L]} p_\ell \beta_\ell \ast \xi_\ell \right) : \xi \in \Delta_L^n \right\}$$

for $t = t^*_k + 1, \ldots, t^*_{k+1}$ do // implement the updated fluid policy

observe incoming supply with utility vector $b_t$;

if $b_t = \beta_\ell$ then

act $x_t \leftarrow (\xi^F)_\ell$;

update $B_{i_t} \leftarrow B_{i_{t-1}} + b_t x_t$ for every agent $i \in [n]$;

**Output:** actions $(x_t : t = 1, 2, \ldots, T)$ and objective $\min_{i \in [n]} B_T^i$. 

Algorithm 4: Frequent Re-solving (FR)

**Input:** time horizon length $T \in \mathbb{N}$, initial utilities $B_0 \in \mathbb{R}^n$.

**Initialize:**

for $t = 1, \ldots, T$ do

- solve the fluid problem with updated utilities:

  $$\xi^F \in \arg \max \left\{ w \left( B_{t-1} + (T - t + 1) \sum_{\ell \in [L]} p_\ell b_\ell * \xi_\ell \right) : \xi \in \Delta^L_n \right\}$$

  observe incoming supply with utility vector $b_t$;

  if $b_i = \beta_\ell$ then
    act $x_t \leftarrow (\xi^F)_\ell$;

  update $B^i_t \leftarrow B^i_{t-1} + b_i x^i_t$ for every agent $i \in [n]$;

**Output:** actions $(x_t : t = 1, 2, \ldots, T)$ and objective $\min_{i \in [n]} B^i_T$. 