Sequential Fair Allocation: Achieving the Optimal Envy-Efficiency Tradeoff Curve

Sean R. Sinclair
srs429@cornell.edu
Cornell University
Ithaca, NY, USA

Siddhartha Banerjee
sbanerjee@cornell.edu
Cornell University
Ithaca, NY, USA

Christina Lee Yu
cleeyu@cornell.edu
Cornell University
Ithaca, NY, USA

ABSTRACT

We consider the problem of dividing limited resources to individuals arriving over T rounds. Each round has a random number of individuals arrive, and individuals can be characterized by their type (i.e. preferences over the different resources). A standard notion of 'fairness' in this setting is that an allocation simultaneously satisfy envy-freeness and efficiency. For divisible resources, when the number of individuals of each type are known upfront, the above desiderata are simultaneously achievable for a large class of utility functions. However, in an online setting when the number of individuals of each type are only revealed round by round, no policy can guarantee these desiderata simultaneously.

We show that in the online setting, the two desired properties (envy-freeness and efficiency) are in direct contention, in that any algorithm achieving additive counterfactual envy-freeness up to a factor of L_T necessarily suffers a efficiency loss of at least $1/L_T$. We complement this uncertainty principle with a simple algorithm, Guarded-Hope, which allocates resources based on an adaptive threshold policy and is able to achieve any fairness-efficiency point on this frontier.

ACM Reference Format:

Sean R. Sinclair, Siddhartha Banerjee, and Christina Lee Yu. 2022. Sequential Fair Allocation: Achieving the Optimal Envy-Efficiency Tradeoff Curve. In *Abstract Proceedings of the 2022 ACM SIGMETRICS/IFIP PERFORMANCE Joint International Conference on Measurement and Modeling of Computer Systems (SIGMETRICS/PERFORMANCE '22 Abstracts), June 6–10, 2022, Mumbai, India.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3489048.3526951

1 INTRODUCTION AND MODEL

Our work here is motivated by a problem faced by a collaborating food-bank (Food Bank for the Southern Tier of New York (FBST)) in operating their mobile food pantry program. In these systems, the mobile food pantry must decide on how much food to allocate to a distribution center on arrival without knowledge of demands in future locations. As a simplified example, every day the mobile food pantry uses a truck to deliver B units of food supplies to individuals over T rounds (where each round can be thought of as a distribution location: soup kitchens, pantries, nursing homes, etc). When the truck arrives at a site t (or round t), the operator observes N_t individuals and chooses how much to allocate to each individual ($X_t \in \mathbb{R}^{N_t}$) before moving to the next round. The number of people assembling at each site changes from day to day, and the operator typically does not know the number of individuals at later sites (but has a sense of the distribution based on previous visits).

In *offline* problems, where the number of individuals at each round $(N_t)_{t\in[T]}$ are known to the principal in advance, there are many well-studied notions of fair allocations of resources. Envy-freeness requires that each individual prefers their own allocation over the allocation of any other. Efficiency requires that the allocations clear the available resources. For divisible resources, the above desiderata are simultaneously achievable for a large class of utility functions, with multiple resources, and is easily computed by maximizing the Nash Social Welfare (NSW) objective subject to allocation constraints. In this (simplified) setting, the fair allocation is easily computed by allocating $X^{opt} = \frac{B}{N}$ to each individual, where $N = \sum_{t\in[T]} N_t$ is the total number of people across all rounds.

Many practical settings, however, the principal makes allocation decisions *online* with incomplete knowledge of the demand for future locations. These principals do have access to historical data allowing them to generate histograms over the number of individuals for each round (or potentially just first moment information). Satisfying any one of these properties is trivially achievable in online settings. The solution that allocates $X_t = 0$ to each individual at location t satisfies hindsight envy-freeness as each individual is given an equal allocation. The solution that allocates $X_1 = B/N_1$ to individuals at the first location and $X_t = 0$ for $t \geq 2$ satisfies efficiency as the entire budget is exhausted at the first location. A more difficult challenge in this setting is achieving low counterfactual envy, ensuring that the allocations made by the algorithm (X_t) are close to what each individual *should* have received with the fair solution in hindsight (B/N).

2 APPROXIMATE FAIRNESS

In sequential settings, one way to measure the *(un) fairness* of any *online* allocation (X^{alg}) is in terms of its counterfactual distance (for both envy and efficiency) when compared to the *optimal fair allocation in hindsight* (i.e., *offline* allocation X^{opt}). Another measure is hindsight envy (when compared only to allocations made by the algorithm). In particular, we define the *counterfactual envy* as $\Delta_{EF} = \|u(X_\theta^{opt}, \theta) - u(X_\theta^{alg}, \theta)\|_\infty$ to be the maximum difference in utility between the algorithm's allocation and the offline allocation where agents are characterized by their type θ, define the *hindsight envy* as $\text{Envy} = \max_{t,t',\theta,\theta'} u(X_{t',\theta'}^{alg}, \theta) - u(X_{t,\theta}^{alg}, \theta)$ to be the maximum difference between the utility individuals would have received if given someone else's allocations, and let $\Delta_{efficiency} =$

The full version of the paper is available at https://arxiv.org/abs/2105.05308.

SIGMETRICS/PERFORMANCE '22 Abstracts, June 6–10, 2022, Mumbai, India

ACM ISBN 978-1-4503-9141-2/22/06.
https://doi.org/10.1145/3489048.3526951

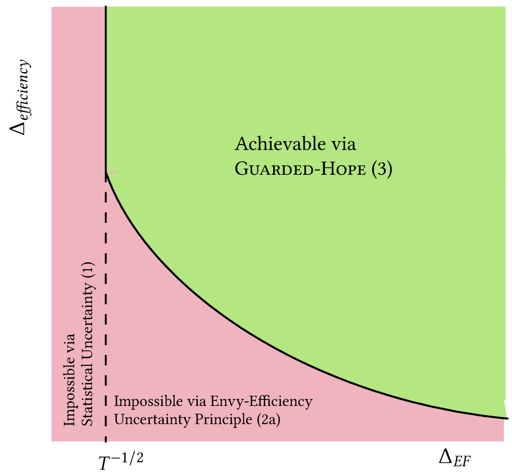

Figure 1: Graphical representation of the major contribution (Informal Theorem 1). The x-axis denotes Δ_{EF} (the maximum difference between utility individuals receive from the algorithm and the fair allocation in hindsight), and the y-axis denotes $\Delta_{efficiency}$, the remaining resources. The dotted line represents the impossibility due to statistical uncertainty in the optimal allocation, and the region below the solid line represents the impossibility due to the envy-efficiency uncertainty principle.

$B - \sum_t N_t X_t^{alg}$ be the algorithm's total leftover resources. These are all very stringent metrics, akin to the notion of regret in online decision-making settings, which subsume many other objectives.

In these settings with competing objectives, practitioners often resort to ad-hoc rules of thumb and trial-and-error adjustments of the system to attempt to manage the balance between objectives. How these criteria interact and trade-off amongst one another is often not well understood or characterized, and furthermore there typically does not exist a clear objective function that determines which tradeoffs are better than others.

3 MAIN RESULTS

Our main technical contribution is to provide a complete characterization of the achievable pairs of (Δ_{EF}, Envy, $\Delta_{efficiency}$). Our results hold in expectation and with high probability, under multiple divisible resources, and with a finite set of individual types with linear utilities. In particular, we show the following informal theorem (see Fig. 1 for a graphical representation).

Informal Theorem 1. *Under mild regularity conditions on the distribution of N_t, we have the following (where $\gtrsim$ ignores problem dependent constants, logarithmic factors of T, and $o(1)$ factors):*

1. (Statistical Uncertainty Principle)*: Any online allocation algorithm must suffer counterfactual envy of at least* $\Delta_{EF} \gtrsim \frac{1}{\sqrt{T}}$.

2a. (Counterfactual Envy-Efficiency Uncertainty Principle)*: Any online allocation algorithm necessarily suffers*

$$\Delta_{efficiency} \gtrsim \min\{\sqrt{T}, 1/\Delta_{EF}\}.$$

2b. (Hindsight Envy-Efficiency Uncertainty Principle)*: Any online allocation algorithm necessarily suffers*

$$\Delta_{efficiency} \gtrsim \min\{\sqrt{T}, 1/\text{Envy}\}.$$

3 (Upper Bound via Guarded-Hope)*: For any choice of L_T, with probability at least $1 - \delta$, Guarded-Hope with parameter L_T achieves:*

$$\text{Envy} \leq L_T \qquad \Delta_{EF} \lesssim \max\{1/\sqrt{T}, L_T\} \qquad \Delta_{efficiency} \lesssim \min\{\sqrt{T}, 1/L_T\}.$$

Furthermore, we provide a simple algorithm, Guarded-Hope, which achieves the correct trade-off between envy and waste, matching the lower bound in terms of T up to logarithmic factors. Our algorithm achieves this using novel concentration arguments on the optimal Nash Social Welfare solution, utilizing a sensitivity argument on the *solution* to the optimization problem instead of the objective (as commonly used for competitive ratio guarantees) to learn a lower guardrail on the optimal solution in hindsight. Given this, we construct an upper guardrail to satisfy the desired Δ_{EF} and Envy bound. We then achieve the proper trade-off by carefully balancing allocating to the established lower guardrail with the upper guardrail while simultaneously ensuring the algorithm never runs out of budget.

To get some intuition into the envy-efficiency uncertainty principle, consider the simple example described above for a single resource (with arrivals N_t in each location, and $X^{opt} = B/N$ where $N = \sum_{t \in [T]} N_t$). For convenience we assume that each agents utility is directly proportional to their allocation (i.e. $u(X, \theta) = X$)). Consider allocation X_1 at the first location: via standard concentration arguments, one can find a high probability lower confidence bound for B/N with a half-width on the order of $1/\sqrt{T}$. Now it's not hard to argue that allocating according to the lower confidence bound at *all* locations achieves counterfactual envy of $\Delta_{EF} \approx 1/\sqrt{T}$, Envy $= 0$, and $\Delta_{efficiency} \approx \sqrt{T}$. This corresponds to the cusp of the efficiency-envy trade-off curves in Fig. 1.

Now if we relax the Δ_{EF} or Envy constraint to $\approx 1/T^{1/3}$ and use the naive static policy of always allocating via the lower confidence bound on that order, we get a waste of $T \cdot T^{-1/3} = T^{2/3}$. Our algorithm instead takes a different approach, using the lower confidence bound of order $1/\sqrt{T}$ as the lower guardrail allocation, and sets the upper guardrail allocation to be the lower one plus the desired bound on Δ_{EF} or Envy. If we were to establish that the algorithm always allocates within the guardrails, we automatically have the desired bound on Δ_{EF} and Envy. The main additional factor in achieving the tradeoff for $\Delta_{efficiency}$ is ensuring we properly allocate according to the upper threshold while ensuring we do not run out of budget to ensure the lower threshold allocation. With this Guarded-Hope achieves $\Delta_{efficiency} \approx T^{1/3}$, which furthermore is the best possible. Moreover, we complement our theoretical results with experiments highlighting the empirical performance of different algorithms (both on synthetic settings, as well as a dataset based on mobile food pantry operations), which shows that Guarded-Hope has much lower waste and envy compared to static under-allocation, as well as other natural heuristics.