Fairly Dividing Non-identical Random Items: Just Sample or Match

Aprup Kale[1], **Navya Garg**[2], **Rucha Kulkarni**[1]
[1]National University of Singapore
[2]Indian Institute of Technology, Bombay

Abstract

We study the question of existence and fast computation of fair and efficient allocations of indivisible resources among agents with additive valuations. As such allocations may not exist for arbitrary instances, we ask if they exist for *typical* or *random* instances, meaning when the utility values of agents for the resources are drawn from certain distributions. In this paper, we extend the previously studied formal models of this problem to non-identical items. We assume that every item is associated with a distribution $\mathcal{U}_j$, and every agent's utility value for the item is drawn independently from $\mathcal{U}_j$. We show that envy-free fair and maximum social welfare efficient allocations exist with high probability in the asymptotic setting, meaning when the number of agents n and items m are large. Further we show that when $m = \Omega(n \log n)$, then by only sampling $O(\log m)$ or $O((\log m)^2)$ utility values per item instead of all the n, we can compute these allocations in $\tilde{O}(m)$ time. Finally, we simulate our algorithms on randomly generated instances and show that even for small instances, we suffer small multiplicative losses in the fairness and efficiency guarantees and converge to fully optimal guarantees quickly.

1 Introduction

Fair and efficient allocation of indivisible resources is a central area at the intersection of economics, algorithmic game theory and computational social choice ([Amanatidis *et al.*, 2023] surveys the most popular directions and seminal results). In this paper, we focus on the particular problem of dividing *typical* or *random* instances (initiated in [Dickerson *et al.*, 2014]), where agent utilities for the resources are drawn from certain distributions rather than being arbitrary. We assume that we are given oracle access to a distribution $\mathcal{U}_j$ (that has minimal technical assumptions described in Section 1.1) corresponding to every item j. Every agent's value for item j is sampled independently from its distribution $\mathcal{U}_j$. This model reasonably captures several real world allocation scenarios. For example, while distributing assets among companies, every company's value for a particular asset would be similar, and can be assumed to belong to some natural distribution with bounded upper and lower values. The same is true for resources like public housing, rent, or time to serve online server requests. If we can guarantee, with high probability, the existence of fair and economically efficient allocations for random instances drawn in this way, and further give computationally efficient algorithms to find these, then we would essentially be able to quickly resolve such real world fair allocation problems with high probability. As there are simple counterexamples to show that fair and efficient allocations may not always exist (consider 2 items and 1 agent), existence with high probability is a strong guarantee. This implication is a compelling reason to pursue this direction of research and the motivation for this paper as well.

The setting of random instances was initiated in [Dickerson *et al.*, 2014]. They studied the formal model of identical goods, where the distributions for the utilities of all the items are identical. In the initial work, they proved the existence of envy-free fair allocations when the number of items m is asymptotically large compared to the number of agents n, that is $m = \Omega(n \log n)$. Their work was followed up by [Manurangsi and Suksompong, 2020] and [Manurangsi and Suksompong, 2021] for the same setting of identical distributions. They closed the gaps in our understanding this setting, and proved tight existence results for fair allocations.

In this paper, we extend this series of works by modelling the class of non-identical distributions, where every item is associated with its own distinct distribution, and show several existence results. First, we show that envy-free fair allocations exist when the number of items m is larger than the number of agents n, that is, $m = \Omega(n \log n)$ (Theorem 1.1). For small m, when $m = O(n)$, we show envy-fair allocations exist when $m \geq 5n$ and n divides m (Theorem 3.2), and proportional allocations exist when $n \leq m \leq 2n$ and the means of all the item distributions are at most $1/2$ (Theorem 3.1), or when $m \geq cn$ for some constant c, that depends on an upper bound of the means of all the item distributions (Theorem 1.4). The overarching techniques to prove these results are similar as the identical case, but the analysis is far more involved. The assumption of identical distributions provides strong symmetry assumptions that are critically used in prior work. The most similar and straightforward is the case for large $m = \Omega(n \log n)$. As in the identical case, we show that there is a gap between the expected utility of an item to an agent who gets the item and that to an agent who does not get it. These values are now different for each item in our non-identical case. We can easily show that there is a gap between the means of these expected values across all items. We then essentially perform the same analysis as the identical case in [Dickerson *et al.*, 2014], by replacing the identical expected values in that analysis by these means of the non-identical expected values.

The linear regime $m = \Theta(n)$ is more involved. In the identical items setting, [Manurangsi and Suksompong, 2020; Manurangsi and Suksompong, 2021] reduce the existence problem of envy-free allocations to the existence of certain many-to-one matchings in certain random bipartite graphs called the Erdős-Rényi graphs. In these graphs, every edge exists with the same probability. Standard known results from combinatorics show when such matchings exist in these graphs. The identical items fair allocation case reduces to this exact problem, hence the known results directly prove the envy-free results. Our non-identical model reduces to a random graph where edges correspond to non-identical item-dependent existence probabilities, and is no longer standard Erdős-Rényi. After constructing this graph, our main technical step is to extend the standard combinatorics result to this heterogenous setting. This yields our envy-free fairness results for linear $m \geq 5n$.

Now we overview our proportionality results. For the $m \geq cn$ case, we follow a similar framework as [Manurangsi and Suksompong, 2021]. First, by a simple pigeonhole principle argument, we prove the sufficiency of the condition on the means of the distributions. For the case when this condition is satisfied, we design a two-stage matching algorithm to constructively prove that a proportional allocation exists. In the first stage, the algorithm constructs a bipartite graph that connects every agent i with any good j only if $u_i(j) \geq \tau_j$, where τ_j is carefully chosen so that $\Pr\left[u_i(j) \geq \tau_j\right]$ is exactly ${\log n + \omega(1)}/{n}$. Our extended proofs for the Erdős-Rényi graphs then imply the existence of certain many-to-one matchings. Now, for the case where $m > 1.9n$, we show by simple concentration bounds that this matching is already proportional. When $m \leq 1.9n$, we need a tighter analysis. This is the most technically involved part of our paper. The difficulty of this case is intuitive, as when m is large, every agent gets a large number of items. Even with non-identical items, the process of combining a large number of them provides some identical measures, like the mean of the expected utility across all items. These measures can be placeholders for the metrics used in the identical items analysis, and make similar proofs go through. We prove a stronger version of the standard Berry-Esseen theorem, extending it to independent but not necessarily identically distributed summands; using this, we show that at most $0.6n$ agents remain without a proportional share after the first stage. For these agents, we construct a residual graph with the remaining agents and items and use this to complete a proportional allocation.

Other works on the random instances model inspired us to investigate our model in two further ways. [Bai and Gölz, 2022] extended research on the random model to the asymmetric agents case, where each agent was associated with a distinct distribution and drew utilities for all the items independently from their own distribution. They also additionally explore Pareto optimality, a notion of economic efficiency. We also investigated economic efficiency, and observed that our algorithm for large number of items $m = \Omega(n \log n)$ case assigns every item to an agent who draws the highest value for it, hence gives Pareto optimality by design. Second, [Manurangsi and Suksompong, 2021] study efficient computation of these allocations. They show that an envy-free allocation can be obtained with a round robin algorithm for $m = \Omega(n \log n / \log \log n)$. [Li *et al.*, 2024] perform a detailed analysis of the round robin algorithm, and show that it has complexity $O(mn + m \log m)$ when agent preferences are uniform, identical and random. These works inspired us to study efficient computation. This seems especially important as the existence results are for the asymptotic regime. Our algorithm for the $m = \Omega(n \log n)$ case has $O(mn)$ time complexity, as for every one of the m items, we draw from their distributions n times, to decide which agent receives it. Note that there is a natural lower bound of $O(m)$ for this problem, as every item needs to be processed once to decide which agent it gets assigned to. We explore if we can match this lower bound. We considered the natural extension to sampling algorithms, where instead of drawing all the n values, we sample a few agents and give the item to the agent who drew the highest utility among these. We show three trade-off results between time and fairness guarantees: by sampling $\tilde{O}(m)$ samples, we obtain 0.8-approximate fair and 0.9-approximately efficient allocations, with high probability (Theorem 1.3). These guarantees can be improved for the more restricted class of discrete distributions (Theorem 1.2), or by drawing more samples (Theorem 4.1).

We now describe our final set of observations. We empirically evaluated both the non-sampling and sampling algorithm on finite instances to understand how quickly the asymptotic behavior manifests. Note that our proofs for the linear $m = O(n)$ case are non-constructive, hence there is no method to experimentally simulate these, hence we simulate our algorithm used in the $m = \Omega(n \log n)$ case. For various values of m and for $n = 100$, we sampled utilities from either the beta, truncated normal, or uniform distributions. The details of our set up are in Section 5. For each instance, we run our algorithm, and record three parameters, (i) envy ratios that capture how much any agent values someone else's set compared to their own, (ii) the fraction of agents who have envy, and (iii) the highest welfare ratio, to understand how much we sacrifice in the total efficiency, to reduce time complexity by sampling. We observed that the envy ratio remains between 4 and 9 when m is small compared to n, but vanishes when m is about 100 times higher than n. We also found that in practice, after sampling about $5 \ln n$ values for every item, the sampling and non-sampling algorithms become nearly indistinguishable in terms of the allocations they yield. Finally, the economic efficiency with the sampling algorithm remains higher than 0.85 times that of the non-sampling one, even when we sample as low as 4 samples per item, and goes to above 0.95 times for 23 or more samples.

Other related work. There are several other works that study the random model. The case of identical undesirable resources or chores is studied by [Manurangsi and Suksompong, 2025]. They showed stronger positive results and proved that the chores case was easier than goods. [Manurangsi *et al.*, 2025] explore a general setting with random identical distributions for utilities, but with weighted agents where each agent has a scalar value called a weight assigned to them. In this problem, fairness notions are generalized to capture the idea that agents receive utilities proportional to their weights. Related works studying similar models are [Bai *et al.*, 2022], which studies fairness in an interesting *smoothed utilities* model, and [Zeng and Psomas, 2020], which studies fair and efficient allocations in online settings and asymmetric agents, and give non-asymptotic results. A series of works studies other notions of fairness with identical items and symmetric agents. [Amanatidis *et al.*, 2017] study maximin share fairness with uniform distributions, [Suksompong, 2016] study proportionality, and [Manurangsi and Suksompong, 2017] study group envy-free fairness, when items are assigned to groups of agents instead of individually. [Kurokawa *et al.*, 2016] studied this problem for maximin share fairness in the more general case of asymmetric agents. [Farhadi *et al.*, 2019] study maximin fairness in the setting with asymmetric agent distributions and weighted agents.

In the remaining paper, we discuss our main results for each case, large and small number of goods - Theorem 3.1 for a proportional allocation when there are $n \leq m \leq 2n$ goods in Section 3, Theorem 3.2 for envy-free allocations with small number of goods in Section 3.2, and Theorem 4.1, our sampling algorithm for large number of items with continuous distributions and bounded means in Section 4. We also describe our empirical evaluations in Section 5. An extended version of this paper with the remaining results can be accessed at: https://arxiv.org/abs/2512.17238.

Theorem 1.1. *[Large number of goods] When the number of goods is large, that is, $n = O(m/\log m)$, then, with probability $(1 - 1/m)$, an envy-free and maximum social welfare allocation exists as $m \to \infty$.*

We write $n = O(m/\log m)$ and $m \to \infty$ to emphasize that the result holds even with a finite number of agents as long as the number of items goes to infinity.

Theorem 1.2. *[Sampling from discrete distributions] Suppose the utility of each good for every agent is drawn from a discrete distribution with finite support in $[0, 1]$. Let $m = \Omega(n \log n)$ and the number of sampled agents for each good $s = \frac{2 \log m}{\alpha_{min}}$. Then, with high probability, a maximum social welfare and envy-free online allocation of goods exists as $n \to \infty$.*

Theorem 1.3. *[Sampling with constant approximation] For the allocation problem as described in Section 1.1, let $m = \Omega(n \log n)$ and the number of sampled agents for each good $s = \frac{20 \log m}{\alpha_{min}}$. Then, with high probability, a 0.9-approx. maximum social welfare and 0.8-approx. envy-free online allocation of goods exists as $n \to \infty$.*

Theorem 1.4. *[Proportionality for linearly many goods] Let $c < 1$ be such that $\mathbb{E}[\mathcal{U}_j] \leq c$ for all $j \in M$. If $r = \lceil 2 \cdot \frac{3+c}{1-c} \rceil$ and $m \geq rn$, then with high probability, a proportional allocation exists as $n \to \infty$.*

1.1 Our Model

Fair allocation problem. We consider the problem of dividing m indivisible and non-identical items among n agents, according to certain fairness and efficiency notions (defined shortly). We assume the items are *goods,* meaning they are non-negatively valued by all the agents. Formally, let $N = \{1, \ldots, n\}$ be the set of agents and $M = \{1, \ldots, m\}$ be the set of goods. For each $(i, j) \in (N \times M)$, $u_i(j) \in [0, 1]$ denotes the *utility* of good j for agent i. The utilities are assumed to be *additive*, that is, for all $M' \subseteq M$, $u_i(M') = \sum_{j \in M'} u_i(j)$. The solution of the fair allocation problem is an allocation or a partition of M into n disjoint subsets. Formally, we denote an allocation by $A = (A_1, \ldots, A_n)$, where A_i is the bundle or subset of M allocated to agent i. Thus $\cup_i A_i = M$ and $A_i \cap A_{i'} = \phi$ for all i and $i' \in N$. In this paper, we will study the setting where each utility $u_i(j)$ is sampled randomly from a known distribution $\mathcal{U}_j$, described below.

Distributions of utilities. For each good $j \in M$, we assume that there exists a distribution $\mathcal{U}_j$ supported on $[0, 1]$, such that $u_i(j) \sim \mathcal{U}_j$ independently for each $i \in N$. Computationally, we assume a black box for each $\mathcal{U}_j$ that allows us to sample once from the distribution in $O(1)$ time. The distributions have the following properties. Each $\mathcal{U}_j$ is *non-atomic*, i.e., $\Pr_{X \sim \mathcal{U}_j}[X = x] = 0$ for all $x \in [0, 1]$. Further, every $\mathcal{U}_j$ is (α_j, β_j)-PDF bounded, meaning the density function of each $\mathcal{U}_j$ has value between α_j and β_j at every point in its support, for some constants $\alpha_j, \beta_j > 0$. Here the α_j and β_j may be distinct for each $j \in M$.

Fairness and efficiency notions. We study the existence and computability of the following three types of allocations. An allocation A is *envy-free* if for all pairs of agents $i, i' \in N$, it holds that $u_i(A_i) \geq u_i(A_{i'})$. The envy of any agent i towards another agent i' is $\max\{0, u_i(A_{i'}) - u_i(A_i)\}$. An allocation A is *proportional* if for all agents $i \in N$, it holds that $u_i(A_i) \geq u_i(M)/n$. Finally, an allocation A has *maximum social welfare* if it maximizes the sum of agent values for their own shares. Formally, if $\mathcal{A}$ is the set of all allocations of M, then $A \in argmax_{\mathcal{A}} \sum_i u_i(A_i)$. We also consider equivalent approximation notions for these. An allocation A is *c-approximate envy-free* ($c < 1$) if for all pairs of agents $i, i' \in N$, it holds that $u_i(A_i) \geq c \cdot u_i(A_{i'})$. Finally, an allocation A has *c-approximate maximum social welfare* ($c < 1$) if some allocation A' has maximum social welfare and $\sum_i u_i(A_i) \geq c \cdot \sum_i u_i(A'_i)$.

2 Matchings in Random Bipartite Graphs

We now discuss the random graphs and their properties used in analyzing the case of small number of items.

Let $\mathcal{G}(n_L, n_R, p)$ denote the distribution of random bipartite graphs with n_L left vertices, n_R right vertices, and p being the probability that there exists an edge between any pair of left and right vertices. A graph sampled from this distribution is called an *Erdős-Rényi random bipartite graph*. The following result is classical and well-known.

Lemma 2.1. *[Erdős and Renyi, 1964]. Let $G \sim \mathcal{G}(n, n, p)$ such that $p = {}^{(\log n + \omega(1))}/_{n}$. Then, with high probability, G contains a perfect matching.*

We prove an extension to Lemma 2.1 where the probability of edges depends on the right vertices.

Theorem 2.1. *Let $G \sim \mathcal{G}(n, n, \{p_j\}_{j\in[n]})$ be a random bipartite graph, say $G = (\{l_i\}_{i\in[n]}, \{r_j\}_{j\in[n]}, E)$. Each edge $(l_i, r_j) \in E$ independently with probability p_j. Define $p_{\min} := \min_{j\in[n]} p_j$. If $p_{\min} \geq {}^{(\log n+\omega(1))}/_{n}$, then with high probability G contains a perfect matching.*

Proof. Given G, define a new random bipartite graph $G' = (\{l_i\}_{i\in[n]}, \{r_i\}_{i\in[n]}, E')$ as follows. If $(l_i, r_j) \in E$, then keep the edge (l_i, r_j) in G' with probability ${}^{p_{min}}/_{p_j}$, and remove it with probability $1 - {}^{p_{min}}/_{p_j}$. Observe that for each (l_i, r_j), $\Pr[(l_i, r_j) \in E'] = \Pr[(l_i, r_j) \in E] \Pr[(l_i, r_j) \in E' \mid (l_i, r_j) \in E] = p_j \cdot {}^{p_{min}}/_{p_j} = p_{min}$.

Thus, G' is a random bipartite graph where each edge is included independently with constant probability $p_{\min}$. Since $p_{\min} \geq {}^{(\log n+\omega(1))}/_{n}$, from Lemma 2.1, G' contains a perfect matching with high probability. Further, $E' \subseteq E$, so G also has a perfect matching with high probability. $\square$

Next, we state the following known result, which proves some sufficient conditions that random graphs must have to satisfy Hall's condition for perfect matchings.

Lemma 2.2. *[Manurangsi and Suksompong, 2021]. Let $G = (A, B, E)$ be a graph sampled from the Erdős-Rényi random bipartite graph distribution $\mathcal{G}(n, q, p)$, where $p \geq 0.5$ and $0.9n \leq q \leq n$. Then with high probability, for every $S \subseteq A$ of size at most $0.6n$, we have $|Z_G(S)| \geq |S|$, where $Z_G(S)$ is the set of vertices adjacent to S in graph G.*

Finally, we talk about right saturated r-matchings.

Definition 2.1. *An r-matching of a bipartite graph $G = (L, R, E)$ is a subset of edges $E' \subseteq E$ such that the subgraph $G' = (L, R, E')$ has degree at most r for vertices in L and degree at most 1 for vertices in R.*

Definition 2.2. *A **right-saturated** r-matching of a bipartite graph $G = (L, R, E)$ is a subset of edges $E' \subseteq E$ such that the subgraph $G' = (L, R, E')$ has degree at most r for vertices in L and degree **exactly** 1 for vertices in R. It is said to be **perfect** if the subgraph G' has degree **exactly** r for vertices in L and degree **exactly** 1 for vertices in R*

Lemma 2.3. *Let $G \sim \mathcal{G}(n, m, p)$ such that $p = {}^{(\log n+\omega(1))}/_{n}$ and $m = rn$. Then, with high probability, G contains a perfect r-matching.*

A simple graph construction, where the new graph contains r copies of all the left vertices in G, followed by showing that each set of left vertices forms a perfect matching with the right vertices, suffices to prove the lemma. A similar reduction, by adding $n - m$ dummy right vertices and edges to these independently with probability p, helps to prove the following.

Lemma 2.4. *Let $G \sim \mathcal{G}(n, m, p)$ such that $p = {}^{(\log n+\omega(1))}/_{n}$ and $n > m$. Then, with high probability, G contains a right saturated matching.*

Algorithm 1 Proportional Algorithm for $n \leq m \leq 2n$

Input: $N, M, \{u_i\}_{i\in N}$
$M^0 :=$ first n items in M, $M^1 := M \setminus M^0$.
$\tau_j \leftarrow 1 - \frac{1.1 \log n}{\alpha_j n}, \quad \forall j \in M$
$E_{\geq\tau} \leftarrow \{(i, j) \in N \times M^0 \mid u_i(j) \geq \tau_j\}$
if $G_{\geq\tau}(N, M^0, E_{\geq\tau})$ has perfect matching **then**
  $(M_1^0, \ldots, M_n^0) \leftarrow$ any perfect matching of $G_{\geq\tau}$
  $N^{\text{vio}} \leftarrow \{i \in N \mid u_i(M_i^0) < \frac{u_i(M)}{n}\}$
  $E^{\text{fix}} \leftarrow \{(i, j) \in N^{\text{vio}} \times M^1 \mid u_i(j) \geq \frac{u_i(M)}{n} - u_i(M_i^0)\}$
  if $G^{\text{fix}}(N^{\text{vio}}, M^1, E^{\text{fix}})$ has $|N^{\text{vio}}|$ size matching **then**
    $(M_1^1, \ldots, M_n^1) \leftarrow$ any such matching
    return $(M_1^0 \cup M_1^1, \ldots, M_n^0 \cup M_n^1)$
  end if
end if

3 Small Number of Goods

3.1 Proportionality for Small Number of Goods

Theorem 3.1. *[Proportionality for small number of goods] For goods, if $n \leq m \leq 2n$ and for all $j, \mathcal{U}_j$ has mean at most ${}^{1}/_{2}$, then, with high probability, a proportional allocation always exists as $n \to \infty$.*

The assumption that each distribution $\mathcal{U}_j$ has a mean at most $1/2$ is crucial to ensure the existence of a proportional allocation for all $n \leq m \leq 2n$ with high probability. Suppose instead $\mathcal{U}_j$ has mean $1/2 + \varepsilon_j$ for some constant $\varepsilon_j > 0$, and set $m = 2n - 1$. Then, the expected utilities satisfy $\mathbb{E}[u_i(M)] = n\left(1 + \frac{1}{n}\sum_j \varepsilon_j - o(1)\right)$. By standard Chernoff and union bounds, with high probability, all agents simultaneously have $u_i(M) > n$. In such cases, no proportional allocation exists since at least one agent must receive at most one item (by the pigeonhole principle), but each item has value at most 1, contradicting proportionality.

Proof. Consider the graph $G_{\geq\tau} = \{N, M^0, E_{\geq\tau}\}$ defined in the first stage of Algorithm 1 with $\tau_j = 1 - \frac{1.1 \log n}{\alpha_j n}$. Since Theorem 2.1 immediately implies that a perfect matching exists in $G_{\geq\tau}$ with high probability, we will assume that this is the case from now on. Also, using Hoeffding's inequality [Hoeffding, 1963], we can show that

$$u_i(M) \leq \frac{m}{2} + 10\sqrt{m \log n} \tag{1}$$

holds with high probability. Hence, we will also assume this from now on. Now, let $q = m - n$ (so $0 \leq q \leq n$). We make two cases based on the value of q.

Case 1: $q < 0.9n$. For sufficiently large n, we have from equation (1), $\frac{u_i(M)}{n} \leq 0.996$ and for sufficiently large n, $\tau_j > 0.996$. Hence, $\tau_j > \frac{u_i(M)}{n}, \forall j \in M$ for all $i \in N$. Hence, running only the first part of the algorithm gives a proportional allocation, i.e., the partial allocation $(M_1^0, \ldots, M_n^0)$ is already proportional.

Case 2: $q \geq 0.9n$. Define $G^1_{\geq\frac{0.5}{\beta}} = (N, M^1, E^1_{\geq\frac{0.5}{\beta}})$ such that $(i, j) \in E^1_{\geq\frac{0.5}{\beta}}$ if and only if $u_i(j) \geq \frac{0.5}{\beta_j}$. Let $Z_G(S)$

denote the set of vertices adjacent to the vertices of the set S in graph G. We define two events $E1$ and $E2$ as follows. $E1$ occurs when $|N^{\text{vio}}| \leq 0.6n$, and $E2$ occurs when $\forall S \subseteq N$ and $|S| \leq 0.6n$, $\left|Z_{G^1_{\geq \frac{0.5}{\beta}}}(S)\right| \geq |S|$.

We first prove that if $E1$ and $E2$ hold with high probability, then the algorithm outputs a proportional allocation. For sufficiently large n, $\frac{u_i(M)}{n} = 1 + o(1)$ (using equation (1)) and $1 + o(1) < \frac{0.5}{\beta_j} + \tau_j$. Now, if $E1$ and $E2$ hold, due to Hall's Marriage Theorem, there exists a matching from N^{vio} to M^1 that uses all vertices in N^{vio}. And since $\frac{u_i(M)}{n} < \frac{0.5}{\beta_j} + \tau_j$, $\forall i \in N$ and $\forall j \in M$, the matching in $G^1_{\geq \frac{0.5}{\beta}}$ from N^{vio} to M^1 also is a matching in G^{fix}. The algorithm thus outputs a proportional allocation. Thus, proving that $E1$ and $E2$ hold with high probability suffices. We now proceed to show this.

Let f_j denote the probability distribution function of $\mathcal{U}_j$. Since $f_j(x) \leq \beta_j, \forall x \in [0,1]$, $\Pr\left[u_i(j) \geq {}^{0.5}/_{\beta_j}\right] \geq 0.5$. Hence, for $G^1_{\geq \frac{0.5}{\beta}}$, Lemma 2.2, directly implies that $E2$ holds with high probability if $E1$ holds.

For showing $E1$, let $X_j = u_i(j) - \mu_j$. Thus $\mathbb{E}[X_j] = 0$ and $\mathbb{E}[X_j^2] = \sigma_j^2$, where σ_j is the standard deviation of the distribution $\mathcal{U}_j$. Also define $\rho_j = \mathbb{E}\left[|X_j|^3\right]$. Thus, $\mathbb{E}\left[X_j^2\right] = \int_{-\mu_j}^{1-\mu_j} x^2 f_j(x)dx$. Since $\alpha_j \leq f_j(x) \leq \beta_j$ for every $x \in [-\mu_j, 1-\mu_j]$,

$$\alpha_j \int_{-\mu_j}^{1-\mu_j} x^2 dx \;\leq\; \mathbb{E}\left[X_j^2\right] \;\leq\; \beta_j \int_{-\mu_j}^{1-\mu_j} x^2 dx$$

$$\alpha_j \frac{1 - 3\mu_j + 3\mu_j^2}{3} \;\leq\; \mathbb{E}\left[X_j^2\right] \;\leq\; \beta_j \frac{1 - 3\mu_j + 3\mu_j^2}{3}$$

Now, since $\mu_j \in [0, 1/2]$, we have $1/4 \leq 1 - 3\mu_j + 3\mu_j^2 \leq 1$, which means

$$\frac{\alpha_j}{12} \;\leq\; \mathbb{E}\left[X_j^2\right] \;\leq\; \frac{\beta_j}{3} \tag{2}$$

Now, consider the quantity $\frac{\sum_j \rho_j}{\left(\sum_j \sigma_j^2\right)^2}$. By Lyapunov's moment inequality (see [Karr, 1993] Corollary 4.37, p.120), we have $\mathbb{E}[|X_j|^3] \leq \mathbb{E}[X_j^2]^{3/2} = \sqrt{\mathbb{E}[X_j^2]}\,\mathbb{E}[X_j^2]$, so

$$\frac{\sum_j \rho_j}{\left(\sum_j \sigma_j^2\right)^2} \leq \frac{\sum_j \sigma_j^3}{\left(\sum_j \sigma_j^2\right)^2} \leq \frac{\sigma_{max} \sum_j \sigma_j^2}{\left(\sum_j \sigma_j^2\right)^2}$$
$$= \frac{\max \sqrt{\mathbb{E}\left[X_j^2\right]}}{\left(\sum_{j=1}^m \mathbb{E}\left[X_j^2\right]\right)} \leq \frac{\max \sqrt{(\beta_j/3)}}{\sum_{j=1}^m (\alpha_j/12)}$$

where the first inequality follows from Lyapunov's inequality and the third inequality follows from equation (2). Thus

$$\frac{\sum_{j \in M} \rho_j}{\left(\sum_{j \in M} \sigma_j^2\right)^2} \leq \frac{4\sqrt{3\beta_{max}}}{m\,\alpha_{min}} \tag{3}$$

where α_{min} and β_{max} represent $\min_{j \in M}(\alpha_j)$ and $\max_{j \in M}(\beta_j)$ respectively. Now, we know that $\Pr[u_i(M) \leq n\min(\tau_j)] \geq \Pr[\sum_{j=1}^m u_i(j) \leq \frac{m}{2}\min(\tau_j)]$ because $m \leq 2n$. Rearranging the terms gives us $\Pr[\frac{u_i(M)}{n} \leq \min(\tau_j)] \geq \Pr[\sum_{j=1}^m X_j \leq \frac{m}{2}\min(\tau_j) - \sum_{j=1}^m \mu_j]$. Now $\sum \mu_j \leq m/2$, as every distribution $\mathcal{U}_j$ has a mean at most $1/2$, thus,

$$\begin{aligned}&\Pr[u_i(M)/n \leq \min(\tau_j)]\\ &\quad\geq \Pr[X_1 + \cdots + X_m \leq (m/2)\min(\tau_j) - m/2]\\ &\quad= \Pr\left[X_1 + \cdots + X_m \leq -\frac{1.1m\log n}{2n\,\alpha_{min}}\right]\end{aligned}$$

Now, let's define $S_m = \frac{X_1 + \cdots + X_m}{\sqrt{\sigma_1^2 + \cdots + \sigma_m^2}}$ and let F_S denote the cumulative distribution function of S_m. Let $t_n = -\frac{1.1m\log n}{2n\,\alpha_{min}}\frac{1}{\sqrt{\sigma_1^2 + \cdots + \sigma_m^2}}$. Then

$$\Pr\left[\frac{u_i(M)}{n} \leq \min(\tau_j)\right] \geq F_S(t_n) \tag{4}$$

and using equation (2), $t_n \geq -\frac{1.1m\log n}{2n\,\alpha_{min}}\frac{1}{\sqrt{m\alpha_{min}/12}}$ and $t \leq 0$. Note also that t_n tends to 0 as n gets sufficiently large.

Now, by [Esseen, 1942], for independent (not necessarily identically distributed) summands, we can say that, $\sup_x |F_S(x) - \Phi(x)| \leq C_0 \cdot \frac{\sum_j \rho_j}{(\sum_j \sigma_j^2)^{3/2}}$, where Φ represents the PDF of the standard normal distribution. Hence, $F_S(t_n) \geq \Phi(t_n) - C_0 \cdot \frac{\sum_j \rho_j}{(\sum_j \sigma_j^2)^{3/2}}$. Using equation (3), $F_S(t_n) \geq \Phi(t_n) - C_0 \cdot \frac{4\sqrt{3\beta_{max}}}{m\,\alpha_{min}}$. Since, α_{min} and β_{max} are finite quantities, for sufficiently large n (and thus also m), and as t_n tends to 0, the above equation becomes $F_S(t_n) \geq \Phi(t_n) - 0.04 \geq 0.49 - 0.04$. Thus, using equation (4), we get $\Pr\left[\frac{u_i(M)}{n} \leq \min(\tau_j)\right] \geq 0.45$ for all agents $i \in N$, so each agent has at most 0.55 probability of being present in N^{vio}. Hence, using standard concentration inequalities, $|N^{\text{vio}}| \leq 0.6n$ with high probability.

Thus, the algorithm outputs a proportional allocation with high probability in both cases. This concludes our proof. $\square$

3.2 Envy-freeness for Small Number of Goods

Theorem 3.2. *[Envy-freeness for small number of goods] When $m \geq 5n$ and m is divisible by n, then with high probability, an envy-free allocation exists as $n \to \infty$.*

Proof. Let $m = xn$ for some $x \in \mathbb{Z}_{\geq 0}$. Let $\alpha_{min} = \min_{j \in M} \alpha_j$ and $\beta_{max} = \max_{j \in M} \beta_j$.

Construct a bipartite graph $G = (N, M, E)$ where edges are given by:

$$E = \{(i,j) \in N \times M : u_i(j) \geq \tau_j\}$$

where $\tau_j := 1 - \frac{1.1\log n}{\alpha_j n}$ and $\tau := \min_{j \in M} \tau_j = 1 - \frac{1.1\log n}{\alpha_{min} n}$.

Let $\tau' := 1 - \frac{1}{n^{4/x+\varepsilon}}$ for some $\varepsilon > 0$ (say $\epsilon = 0.1$) and $x \geq 5$. The graph G is an Erdős-Rényi random graph with edge probabilities at least $\frac{\log n + \omega(1)}{n}$, so by Lemma 2.3 with high probability, there exists a perfect x-matching corresponding to an allocation $\{A_1, \ldots, A_n\}$ such that:

$$u_i(A_i) \geq x\tau$$

Now fix any pair $i, i' \in N$. The probability that more than $\frac{x}{2}$ goods in A_i have $u_{i'}(j) > \tau'$ is bounded by:

$$\Pr\left(|\{j \in A_i : u_{i'}(j) > \tau'\}| > \frac{x}{2}\right) \leq \left(\frac{\beta_{max}}{n^{4/x+\varepsilon}}\right)^{x/2} = o(n^{-2})$$

By union bound over all $i, i' \in N$:

$$\Pr\left(\exists i, i' \in N : |\{j \in A_i : u_{i'}(j) > \tau'\}| > \frac{x}{2}\right) = n^2 \cdot o(n^{-2}) = o(1)$$

Hence, with high probability:

$$u_i(A_{i'}) \leq \frac{x}{2}\tau' + \frac{x}{2} \Rightarrow u_i(A_i) - u_i(A_{i'}) \geq x\tau - \left(\frac{x}{2}\tau' + \frac{x}{2}\right) = \frac{x}{2}\left(\frac{1}{n^{4/x+\varepsilon}} - \frac{2.2\log n}{\alpha_{min} n}\right)$$

We want this to be at least 0. Thus,

$$\frac{x}{2}\left(\frac{1}{n^{4/x+\varepsilon}} - \frac{2.2\log n}{\alpha_{min} n}\right) \geq 0$$

For sufficiently large n, and x strictly greater than 4, $\frac{1}{n^{4/x+\varepsilon}} \geq \frac{2.2\log n}{\alpha_{min} n}$.
The allocation $(A_1, \ldots, A_n)$ is thus envy-free. □

4 Large Number of Goods

We will assume the sampling model as follows. We assume that for each good $j \in M$, we sample a set $S_j \subset N$ of s agents at random (uniformly). We assign each j to the agent in S_j who draws the highest value for it. We will show that this simple algorithm gives strong envy-free and maximum social welfare guarantees. Our model is as discussed in Section 1.1, with the following additional mild constraint. As long as we are guaranteed a constant upper bound (say $\mu < 1$) on the means of the distributions of the goods, that is, $\mu_j < 1$ for all goods j, we have the following result.

Theorem 4.1. *[Sampling with near optimal guarantees] Let $m = \Omega(n \log n)$ and the number of sampled agents for each good $s = \frac{2(\log m)^2}{\alpha_{min}}$. If, for all goods j, $\mathbb{E}[\mathcal{U}_j] \leq \mu$ for some constant $\mu < 1$, then, with high probability, a $(1 - \frac{1}{\log n})$-approx. maximum social welfare and envy-free online allocation of goods exists as $n \to \infty$.*

Proof. For any good j,

$$\Pr\left(\forall i \in S_j, u_i(j) < 1 - \frac{1}{\log n}\right) \leq \left(1 - \frac{\alpha_j}{\log n}\right)^s$$

By union bound over all goods,

$$\Pr\left(\exists j \text{ s.t. } \forall i \in S_j, u_i(j) < 1 - \frac{1}{\log n}\right) \leq \sum_{j \in M}\left(1 - \frac{\alpha_j}{\log n}\right)^s \leq m\left(1 - \frac{\alpha_{min}}{\log n}\right)^s \leq m \cdot e^{-\frac{\alpha_{min}}{\log n}\cdot s} \leq \frac{1}{m}$$

where the last inequality is due to the chosen value of s. Thus, with high probability, for all goods, there is always at least one agent in the sampled set of agents that has value greater than or equal to $1 - \frac{1}{\log n} \geq \frac{\mu+1}{2}$ (for sufficiently large n). Hence, the allocation is $(1 - \frac{1}{\log n})$- maximum social welfare by construction. If good j is allocated to agent i (that is $j \in A_i$), we also know that with high probability $u_i(j) \geq \frac{\mu+1}{2}$. This in turn implies that, with high probability, $\mathbb{E}\left[u_i(j) | j \in A_i\right] \geq \frac{\mu+1}{2}$. We now show that the allocation is EF with high probability. Define for fixed $i \in N$, $j \in M$:

$$X_i^j = \begin{cases} u_i(j), & \text{if } j \in A_i \\ 0, & \text{otherwise} \end{cases}$$

Then $u_i(A_i) = \sum_{j \in M} X_i^j$. Moreover,

$$\mathbb{E}[X_i^j] = \Pr\left[j \in A_i\right] \cdot \mathbb{E}\left[u_i(j) \,|\, j \in A_i\right] \geq \frac{1}{n} \cdot \frac{\mu+1}{2}$$

Using linearity of expectation,

$$\mathbb{E}[u_i(A_i)] = \sum_{j \in M} \mathbb{E}[X_i^j] \geq \frac{\mu+1}{2} \cdot \frac{m}{n}$$

Now, define for $i, i' \in N$, $i \neq i'$, and $j \in M$:

$$Y_{ii'}^j = \begin{cases} u_i(j), & \text{if } j \in A_{i'} \\ 0, & \text{otherwise} \end{cases}$$

Then $u_i(A_{i'}) = \sum_{j \in M} Y_{ii'}^j$. Furthermore,

$$\mathbb{E}[Y_{ii'}^j] = \frac{1}{n} \cdot \mathbb{E}\left[u_i(j) \,|\, j \in A_{i'}\right] = \frac{1}{n} \cdot \mathbb{E}\left[u_i(j) \,|\, j \notin A_i\right] \leq \frac{\mu}{n}$$

(since $\mathbb{E}\left[u_i(j) \,|\, j \notin A_i\right] \leq \mu_j \leq \mu$, where the first inequality holds from the proof of Theorem 1.2.

Let $Z_{ii'}^j \in [0, 1]$ be such that $\mathbb{E}[Z_{ii'}^j] = \frac{\mu}{n}$, and $Z_{ii'}^j$ stochastically dominates $Y_{ii'}^j$. Then, $\sum_{j \in M} \mathbb{E}[Z_{ii'}^j] = \mu \cdot \frac{m}{n}$. Let $E_{ii'}$ be the event that agent i envies agent i', i.e., $\sum_{j \in M} Y_{ii'}^j > \sum_{j \in M} X_i^j$. This occurs only if either of the two conditions below occur:

$$\sum_{j \in M} X_i^j \leq \frac{3\mu+1}{4} \cdot \frac{m}{n} = \left(1 - \frac{1-\mu}{2(\mu+1)}\right)\frac{\mu+1}{2} \cdot \frac{m}{n} \leq \left(1 - \frac{1-\mu}{2(\mu+1)}\right) \cdot \mathbb{E}\left[\sum_{j \in M} X_i^j\right]$$

$$\sum_{j \in M} Y_{ii'}^j \geq \frac{3\mu+1}{4} \cdot \frac{m}{n} = \left(1 + \frac{1-\mu}{4\mu}\right)\mu \cdot \frac{m}{n} = \left(1 + \frac{1-\mu}{4\mu}\right) \cdot \mathbb{E}\left[\sum_{j \in M} Z_{ii'}^j\right]$$

Let $\epsilon = \frac{1-\mu}{2(\mu+1)} < 1$, so $\epsilon < \frac{1-\mu}{4\mu}$ also holds.

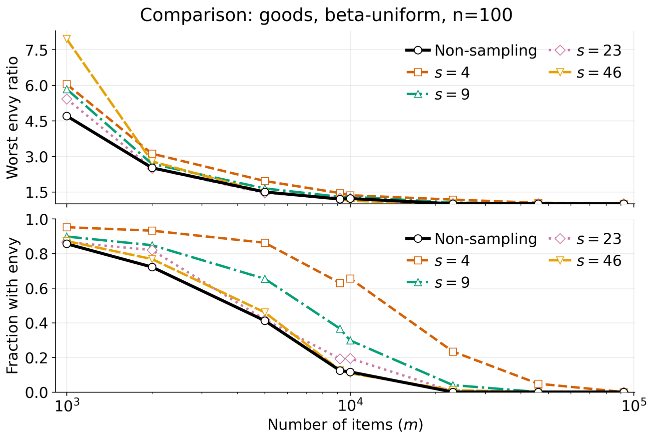

Figure 1: Goods: `beta_uniform` utilities, comparison of sampling and non-sampling algorithms.

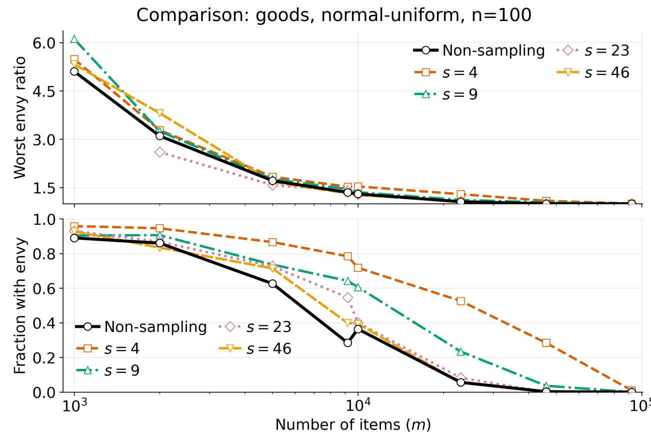

Figure 2: Goods: `normal_uniform` utilities, comparison of sampling and non-sampling algorithms.

Since X_i^j and $Z_{ii'}^j$ are independent across j, using Chernoff bounds:

$$\Pr\left[\sum_{j\in M} X_i^j \le (1-\epsilon)\cdot \mathbb{E}\left[\sum_{j\in M} X_i^j\right]\right] \le \exp\left(-\frac{\epsilon^2}{2}\cdot\frac{\mu+1}{2}\cdot\frac{m}{n}\right)$$

$$\Pr\left[\sum_{j\in M} Y_{ii'}^j \ge (1+\epsilon)\cdot \mathbb{E}\left[\sum_{j\in M} Z_{ii'}^j\right]\right] \le \exp\left(-\frac{\epsilon^2}{3}\cdot\mu\cdot\frac{m}{n}\right)$$

Choose m large enough such that $n \le \frac{\epsilon^2}{3}\cdot\mu\cdot\frac{m}{\ln(2m^3)}$ (since $m = \Omega(n\log n)$). By the union bound:

$$\Pr[E_{ii'}] \le \exp\left(-\frac{\epsilon^2}{2}\cdot\frac{\mu+1}{2}\cdot\frac{m}{n}\right) + \exp\left(-\frac{\epsilon^2}{3}\cdot\mu\cdot\frac{m}{n}\right) \le 2\cdot\frac{1}{2m^3} = \frac{1}{m^3}$$

Allocation A is EF if and only if no event $E_{ii'}$ occurs. The probability that A is not EF is at most:

$$\Pr\left[\bigvee_{\substack{i,i'\in N\\ i\ne i'}} E_{ii'}\right] \le \sum_{\substack{i,i'\in N\\ i\ne i'}} \Pr[E_{ii'}] \le \binom{n}{2}\cdot\frac{1}{m^3} \le \frac{1}{m}$$

Thus, the probability that A is not EF vanishes as $m\to\infty$ ($n\to\infty$ and $m=\Omega(n\log n)$). □

5 Empirical Results

We empirically evaluate both the non-sampling and sampling algorithm on finite instances to understand how quickly the asymptotic behaviour manifests. For each experiment, we fix $n = 100$ agents and vary the number of items m from 10^3 to 10^5. Each item is associated with a **distinct** distribution drawn from one of these classes.

`beta_uniform` mixture. For each item j, and for all agents i, $u_i(j)$ is drawn from the following distributions:

- a Beta(α,β) distribution, with α,β resampled independently for every item, or
- a Uniform$[a,b]$ distribution, with a,b likewise resampled per item.

`normal_uniform` mixture. For each item j, and for all agents i, $u_i(j)$ is drawn from the following distributions:

- a truncated (values are truncated to $[0,1]$ and then normalized so that $\Pr([0,1]) = 1$ for the distribution) Normal(μ,σ^2) value, with (μ,σ) resampled independently per item, or
- a Uniform$[a,b]$ value as above.

Agent utilities are then drawn from the item distributions, and allocations computed as per the maximum welfare producing algorithm. We record three quantities for each allocation: the **worst envy ratio** (the maximum envy relative to their own set's value over all agents), the **fraction of agents with any envy**, and the **welfare ratio** comparing sampled and non-sampled allocations. Each experimental point represents the average of these values over multiple independent trials.

Envy behaviour. Across all distributions, envy drops rapidly as m increases (Figures 1 and 2). Worst envy ratios begin between roughly 4-9 at $m = 10^3$ but approach 1 by about $m = 5\times 10^4$. The fraction of agents with any envy similarly falls to zero as m grows. Sampling introduces the expected trade-off: small s leads to higher envy at moderate m, but as soon as $s \ge 23 \approx 5\ln n$, it becomes nearly indistinguishable from the non-sampling algorithm.

Welfare under sampling. The welfare ratio between sampled and non-sampled allocations remains high throughout Even with $s = 4$, the ratio stays above 0.85, and for $s = 23$ or 46 it consistently exceeds 0.95 across all values of m. The ratio is essentially flat in m, indicating that the efficiency impact depends primarily on s.

From these observations, we can conclude that the sampling algorithm reaches the guarantees of the non-sampling algorithm for large m, thus offers a practical and scalable alternative for large fair division instances.

Acknowledgments

This work was supported by the NRF grant NRF-NRFI09-0005.

References

[Amanatidis *et al.*, 2017] Georgios Amanatidis, Evangelos Markakis, Afshin Nikzad, and Amin Saberi. Approximation algorithms for computing maximin share allocations. *ACM Transactions on Algorithms (TALG)*, 13(4):1–28, 2017.

[Amanatidis *et al.*, 2023] Georgios Amanatidis, Haris Aziz, Georgios Birmpas, Aris Filos-Ratsikas, Bo Li, Hervé Moulin, Alexandros A Voudouris, and Xiaowei Wu. Fair division of indivisible goods: Recent progress and open questions. *Artificial Intelligence*, 322:103965, 2023.

[Bai and Gölz, 2022] Yushi Bai and Paul Gölz. Envy-free and pareto-optimal allocations for agents with asymmetric random valuations. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, IJCAI-2022, page 53–59. International Joint Conferences on Artificial Intelligence Organization, July 2022.

[Bai *et al.*, 2022] Yushi Bai, Uriel Feige, Paul Gölz, and Ariel D Procaccia. Fair allocations for smoothed utilities. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 436–465, 2022.

[Dickerson *et al.*, 2014] John Dickerson, Jonathan Goldman, Jeremy Karp, Ariel Procaccia, and Tuomas Sandholm. The computational rise and fall of fairness. *Proceedings of the AAAI Conference on Artificial Intelligence*, 28(1), Jun. 2014.

[Erdős and Renyi, 1964] Paul Erdős and Alfred Renyi. On random matrices. *Magyar Tud. Akad. Mat. Kutató Int. Közl*, 8:455–461, 1964.

[Esseen, 1942] Carl-Gustav Esseen. On the Liapunoff limit of error in the theory of probability. *Arkiv för Matematik, Astronomi och Fysik*, 28A(9):1–19, 1942.

[Farhadi *et al.*, 2019] Alireza Farhadi, Mohammad Ghodsi, Mohammad Taghi Hajiaghayi, Sebastien Lahaie, David Pennock, Masoud Seddighin, Saeed Seddighin, and Hadi Yami. Fair allocation of indivisible goods to asymmetric agents. *Journal of Artificial Intelligence Research*, 64:1–20, 2019.

[Hoeffding, 1963] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, March 1963.

[Karr, 1993] Alan F. Karr. *Probability*. Springer New York, New York, NY, 1993.

[Kurokawa *et al.*, 2016] David Kurokawa, Ariel Procaccia, and Junxing Wang. When can the maximin share guarantee be guaranteed? In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30, 2016.

[Li *et al.*, 2024] Zihan Li, Pasin Manurangsi, Jonathan Scarlett, and Warut Suksompong. Complexity of round-robin allocation with potentially noisy queries. In *International Symposium on Algorithmic Game Theory*, pages 520–537. Springer, 2024.

[Manurangsi and Suksompong, 2017] Pasin Manurangsi and Warut Suksompong. Asymptotic existence of fair divisions for groups. *Mathematical Social Sciences*, 89:100–108, 2017.

[Manurangsi and Suksompong, 2020] Pasin Manurangsi and Warut Suksompong. When do envy-free allocations exist? *SIAM Journal on Discrete Mathematics*, 34(3):1505–1521, January 2020.

[Manurangsi and Suksompong, 2021] Pasin Manurangsi and Warut Suksompong. Closing gaps in asymptotic fair division. *SIAM Journal on Discrete Mathematics*, 35(2):668–706, 2021.

[Manurangsi and Suksompong, 2025] Pasin Manurangsi and Warut Suksompong. Asymptotic fair division: Chores are easier than goods. In James Kwok, editor, *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 3988–3995. International Joint Conferences on Artificial Intelligence Organization, 8 2025. Main Track.

[Manurangsi *et al.*, 2025] Pasin Manurangsi, Warut Suksompong, and Tomohiko Yokoyama. Asymptotic analysis of weighted fair division. In James Kwok, editor, *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence, IJCAI-25*, pages 3979–3987. International Joint Conferences on Artificial Intelligence Organization, 8 2025. Main Track.

[Suksompong, 2016] Warut Suksompong. Asymptotic existence of proportionally fair allocations. *Mathematical Social Sciences*, 81:62–65, 2016.

[Zeng and Psomas, 2020] David Zeng and Alexandros Psomas. Fairness-efficiency tradeoffs in dynamic fair division. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 911–912, 2020.