

Fair Allocation of Indivisible Goods with Variable Groups

Paul Gözl¹, Ayumi Igarashi², Pasin Manurangsi³, Warut Suksompong⁴

¹Cornell University, USA

²University of Tokyo, Japan

³Google Research, Thailand

⁴National University of Singapore, Singapore

Abstract

We study the fair allocation of indivisible goods with *variable groups*. In this model, the goal is to partition the agents into groups of given sizes and allocate the goods to the groups in a fair manner. We show that for any number of groups and corresponding sizes, there always exists an *envy-free up to one good (EF1)* outcome, thereby generalizing an important result from the individual setting. Our result holds for arbitrary monotonic utilities and comes with an efficient algorithm. We also prove that an EF1 outcome is guaranteed to exist even when the goods lie on a path and each group must receive a connected bundle. In addition, we consider a probabilistic model where the utilities are additive and drawn randomly from a distribution. We show that if there are n agents, the number of goods m is divisible by the number of groups k , and all groups have the same size, then an *envy-free* outcome exists with high probability if $m = \omega(\log n)$, and this bound is tight. On the other hand, if m is not divisible by k , then an envy-free outcome is unlikely to exist as long as $m = o(\sqrt{n})$.

1 Introduction

Fairly allocating limited resources is a fundamental societal challenge, with applications ranging from dividing household supplies among families to distributing personnel among schools or other public institutions. The problem of *fair division* has accordingly received interest in several disciplines, including in computational social choice and multi-agent systems (Bouveret, Chevaleyre, and Maudet 2016; Markakis 2017; Aziz 2020; Walsh 2020).

Most of the work in fair division assumes that each recipient of a bundle of resources is an *individual* agent, represented by a single preference. However, when distributing resources among families, schools, or institutions, each recipient in fact consists of multiple agents. Although these agents share the same set of resources and derive full value from the resources in their set, they may have differing preferences over the resources. This has motivated several researchers to study the *(fixed-)group* model, where the agents are partitioned into groups and the aim is to allocate the resources in a fair manner among the groups (Manurangsi and Suksompong 2017, 2025b; Ghodsi et al. 2018; Segal-Halevi and Nitzan 2019; Kyropoulou, Suksompong, and Voudouris 2020; Segal-Halevi and Suksompong 2023; Caragiannis, Larsen, and Shyam 2025; Gözl and Yaghoubizade

2026; Manurangsi and Meka 2026). For instance, Manurangsi and Suksompong (2022) investigated the notion of *envy-freeness up to c goods (EF c)*. This means that no agent would prefer another group’s bundle over her own group’s if some set of at most c goods were removed from the other group’s bundle. Manurangsi and Suksompong proved that when the number of groups is constant, there exists an EF c allocation for $c = O(\sqrt{n})$, where n denotes the total number of agents, and this bound is also asymptotically tight.

The fixed-group model is appropriate when membership in the groups is predetermined, as in the allocation among families or countries. In other applications, however, the resource allocator can select the partition of agents into groups alongside the allocation. This is the case, for example, when dividing workers in an organization into teams and assigning resources to these teams. In light of this, Kyropoulou, Suksompong, and Voudouris (2020, Sec. 5) proposed a *variable-group* model, in which a partition of the agents into groups can be chosen along with an allocation of the resources.¹

When the resources consist of *divisible* items such as time, Segal-Halevi and Suksompong (2021) proved that an envy-free outcome always exists in the variable-group model. On the other hand, if the resources contain *indivisible* items such as books or gym equipment, Kyropoulou, Suksompong, and Voudouris (2020) showed that an EF1 outcome can be ensured in the case of two groups, and an outcome satisfying a relaxation of proportionality—which is fundamentally weaker than EF1—can be satisfied for any number of groups. This differs from the fixed-group model, where even for two groups, the optimal EF c guarantee deteriorates as the number of agents grows.

In this paper, we expand and deepen our understanding of fairness, particularly envy-freeness, when allocating indivisible goods in the variable-group model. For instance, we study the following question: does an EF1 outcome exist for any number of groups and any corresponding sizes, or is it sometimes necessary to relax the notion to EF c for some (possibly non-constant) c ? As we shall see, the flexibility provided by this model enables remarkably strong fairness guarantees to be made.

¹For further motivation of the variable-group model, we refer to the papers by Kyropoulou, Suksompong, and Voudouris (2020) and Segal-Halevi and Suksompong (2021).

1.1 Our Results

In Section 3, we answer the question above in the positive: for any desired group sizes, an EF1 outcome exists and, moreover, can be computed efficiently (Theorem 1). This result holds even for arbitrary monotonic utilities, and significantly generalizes the well-known result by Lipton et al. (2004) that EF1 allocations always exist in the *individual setting* where each group has size one. Our algorithm is a careful extension of Lipton et al.’s classic *envy cycle elimination algorithm*. This positive result stands in contrast to the fixed-group setting, where $c = \Omega(\sqrt{n})$ is required to guarantee the existence of an EF c allocation (Manurangsi and Suksompong 2022).

In Section 4, we strengthen the previous result by showing that, if the goods lie on a path, EF1 allocations exist even when each group must receive a connected bundle (Theorem 2). The connectivity requirement is natural when the goods correspond to time slots or offices along a corridor, and has been studied in several papers (Bouveret et al. 2017; Bei et al. 2022; Bilò et al. 2022; Igarashi 2023). This existence result also holds for monotonic utilities, and generalizes results from the individual setting by Bilò et al. (2022) and Igarashi (2023). However, like in the individual setting, the result does not come with an efficient algorithm.

Finally, in Section 5, we consider a probabilistic model where the utilities are additive and each agent’s utility for each good is drawn independently at random from a non-atomic distribution.² We are interested in when an *envy-free*³ outcome exists with high probability (that is, with probability approaching 1 as the number of agents n grows), assuming that the number of groups k is fixed and all groups have the same size. Interestingly, we show that the existence depends on whether the number of goods m is divisible by k . On the one hand, if m is divisible by k , the transition occurs at merely $\Theta(\log n)$: an envy-free outcome is unlikely to exist if $m = o(\log n)$ (Theorem 3), but likely to exist if $m = \omega(\log n)$ (Theorem 4). On the other hand, if m is not divisible by k , such an outcome is unlikely to exist as long as $m = o(\sqrt{n})$ (Theorem 5).⁴

1.2 Further Related Work

As mentioned earlier, a number of authors have investigated fair division among groups, mostly focusing on the fixed-group model. Besides envy-freeness, Manurangsi and Suksompong (2022) also obtained bounds on *proportionality* as well as *consensus 1/k-division*—the latter is even more stringent than envy-freeness, as it requires agents to value all bundles of goods equally. While these bounds were already tight in terms of n , Caragiannis, Larsen, and Shyam (2025) and Manurangsi and Meka (2026) recently improved their

dependence on k . Bu et al. (2023), Barman et al. (2025), as well as Kawase, Roy, and Sanpui (2025) studied settings that can be interpreted as special cases of the fixed-group model. For example, Bu et al.’s setting corresponds to the fixed-group model when each group has size two.

The probabilistic model we consider in Section 5 falls under the framework of *asymptotic fair division*. This framework was introduced by Dickerson et al. (2014) and subsequently studied in several papers (Kurokawa, Procaccia, and Wang 2016; Suksompong 2016; Manurangsi and Suksompong 2020, 2021, 2025a; Bai and Gözl 2022; Bai et al. 2022; Benadè et al. 2024; Manurangsi, Suksompong, and Yokoyama 2025; Yokoyama and Igarashi 2025). The motivation is that, since an allocation satisfying envy-freeness (or some other fairness notion) does not always exist, it is natural to ask *when* an envy-free allocation is likely to exist if the utilities are drawn at random. We highlight two relations between our results and known results from this line of work. Firstly, Manurangsi and Suksompong (2017) examined the fixed-group model and showed that an envy-free allocation is unlikely to exist unless $m = \Omega(n)$. This contrasts with our results, which show that existence is already likely in the variable-group model when $m = \omega(\log n)$. Secondly, in the individual setting, Manurangsi and Suksompong (2020, 2021) proved that the threshold for the existence of envy-free allocations differs according to whether m is divisible by n . However, the (multiplicative) gap in their case is only logarithmic (i.e., $\Theta(n)$ vs. $\Theta(n \log n / \log \log n)$), whereas our gap is much larger (i.e., $\Theta(\log n)$ vs. $\Omega(\sqrt{n})$).

2 Preliminaries

For any positive integer t , let $[t] := \{1, 2, \dots, t\}$. Let $k \geq 2$ and $n_1, \dots, n_k$ be positive integers, $N = [n]$ be a set of $n = n_1 + \dots + n_k$ agents, and $M = [m]$ be a set of m goods. A *bundle* refers to a (possibly empty) set of goods. Each agent $a \in N$ has a utility function u_a over the sets of goods in M ; for a single good $g \in M$, we sometimes write $u_a(g)$ instead of $u_a(\{g\})$. We assume that the utilities are *monotonic*, meaning that $u_a(M') \leq u_a(M'')$ for any $a \in N$ and $M' \subseteq M'' \subseteq M$, and *normalized*, i.e., $u_a(\emptyset) = 0$. The utilities are called *additive* if $u_a(M') = \sum_{g \in M'} u_a(g)$. When utilities are non-additive, we assume that an algorithm can query any agent’s utility for any set of goods in constant time. An *instance* in the variable-group model consists of the set of agents N , the set of goods M , the agents’ utility functions, and the desired group sizes $n_1, \dots, n_k$.

We would like to partition the n agents into k groups $C_1, \dots, C_k$ of sizes $n_1, \dots, n_k$, respectively, and allocate the m goods among these groups. We write $C = (C_1, \dots, C_k)$. An *allocation* $A = (A_1, \dots, A_k)$ consists of k disjoint bundles, where bundle A_i is allocated to the i -th group; it is called *complete* if $A_1 \cup \dots \cup A_k = M$. An *outcome* consists of a partition of agents C along with a complete allocation of goods A . An outcome (C, A) is

- *envy-free up to c goods (EF c)*, for a given non-negative integer c , if for every $i, j \in [k]$ and every agent $a \in C_i$, there exists a set $B \subseteq A_j$ with $|B| \leq c$ such that $u_a(A_i) \geq u_a(A_j \setminus B)$;

²A distribution is called *non-atomic* if it does not put a positive probability on any single point.

³Envy-freeness corresponds to EF c for $c = 0$. An envy-free outcome does not always exist, e.g., when there are two groups and only one valuable good.

⁴In contrast, a result in the fixed-group setting by Manurangsi and Suksompong (2017) implies that an envy-free outcome is likely to exist when $m = \Omega(n \log n)$.

- *envy-free* if it is EF0.

The *individual setting* refers to a special case of this model where $n_1 = \dots = n_k = 1$ (and therefore $k = n$).

3 EF1: Existence and Computation

Recall that in the individual setting, a classic result of Lipton et al. (2004) states that for any instance with arbitrary monotonic utilities, an EF1 allocation exists. Moreover, such an allocation can be found in time $O(mn^3)$ via the *envy cycle elimination algorithm*. Intuitively, this algorithm allocates one good at a time in an arbitrary order. The algorithm maintains an “envy graph”, which is a directed graph that captures the envy relations among the agents. In particular, the vertices represent the agents, and there is an edge from one agent to another agent if and only if the former agent envies the latter agent. Hence, an agent is unenvied exactly when the corresponding vertex has no incoming edge. Each good is assigned to an unenvied agent, and if the assignment causes a cycle to be formed in the envy graph, the cycle is eliminated by giving each agent’s bundle to the preceding agent on the cycle. Once all cycles have been eliminated, the next good can be assigned to an unenvied agent, and EF1 is maintained throughout the algorithm.

A priori, it may appear that the envy cycle elimination algorithm is not well-suited for the group setting. Indeed, it is unclear when a group should be considered to “envy” another group, as different agents in the same group may have differing opinions about the groups’ bundles. Nevertheless, we show that adopting an alternative perspective on the algorithm allows for its generalization to the group setting. Specifically, instead of moving the *bundles* as in the typical interpretation, we can reinterpret the envy cycle elimination step as moving the *agents* along the cycles instead. This interpretation enables a generalization of the algorithm to accommodate groups, as it permits the reassignment of individual agents according to their envy relations while preserving group sizes, thereby leveraging the flexibility of the variable-group model. In addition to extending the seminal result of Lipton et al. (2004), the following theorem also strengthens a variable-group guarantee due to Kyropoulou, Suksompong, and Voudouris (2020, Thm. 5.6), which holds under additive utilities for a much weaker notion than EF1.

Theorem 1. *For any instance with arbitrary monotonic utilities, there exists an EF1 outcome. Moreover, such an outcome can be computed in time $O(mn^3)$.*

Proof. We use the following generalization of the envy cycle elimination algorithm.

1. Let C be an arbitrary partition of the n agents into groups of sizes $n_1, \dots, n_k$, and let A be an empty allocation.
2. Construct an *envy graph*, which is a directed graph with the k groups as the vertices; this graph will be updated as the algorithm proceeds. For each agent, add an edge from the agent’s group to another group if the agent envies the latter group. (Thus, the graph initially contains no edges.) In particular, there can be multiple edges from one vertex to another vertex.

3. Take an arbitrary unallocated good, and allocate it to any group with no incoming edge in the envy graph. Update the envy graph.
4. If there is at least one directed cycle in the envy graph, consider an arbitrary cycle. For each edge in the cycle, move the agent associated with this edge to the group that the agent envies. Update the envy graph. If there is still a directed cycle in the envy graph, repeat this step.
5. If there is still an unallocated good, go back to Step 3. Otherwise, output the current outcome (C, A) .

We first show that the algorithm outputs a valid outcome. To this end, we prove that each time we eliminate a cycle in Step 4, the number of edges in the envy graph decreases. For each agent *not* associated with an edge in the cycle, the agent’s own bundle as well as all other bundles remain the same, so the number of envy edges from the agent also remains the same. On the other hand, for each agent associated with an edge in the cycle, the agent is assigned to a better bundle in her view among the k bundles, so the number of envy edges from the agent decreases by at least one. Hence, the total number of envy edges decreases, which means that the process of eliminating cycles must terminate. When the envy graph contains no cycle, there must exist a vertex with no incoming edge, and we can allocate the next good to the corresponding group. It follows that all goods are allocated. Moreover, the sizes of the k groups remain $n_1, \dots, n_k$ throughout, so the algorithm outputs a valid outcome.

Next, we show that the outcome returned by the algorithm is EF1. Specifically, we prove that at every point during the execution of the algorithm, the partition C and the (possibly incomplete) allocation A together yield EF1. This is true at the beginning of the algorithm, as the allocation is empty. When a good is allocated, it is allocated to a group with no incoming edge, so any envy towards the group can be eliminated by removing this good. Moreover, when a cycle is eliminated, for each agent *not* associated with an edge in the cycle, the agent’s own bundle as well as all other bundles remain the same, so the EF1 invariant is maintained. On the other hand, for each agent associated with an edge in the cycle, the agent is assigned to a better bundle in her view among the k bundles. Since any envy that the agent has towards another bundle can be eliminated by removing a good from the bundle before, the same remains true afterwards, and the EF1 invariant is again maintained.

Finally, we analyze the running time of the algorithm. There are m allocated goods, and each allocated good increases the number of envy edges by at most n . Finding and eliminating a directed cycle can be done in time $O(k^2)$, and the elimination decreases the number of envy edges by at least one. As the algorithm can query any agent’s utility for any set of goods in constant time, updating the envy graph takes time $O(nk)$. Since $k \leq n$, the algorithm runs in time $O(mn^3)$. $\square$

4 EF1: Adding Connectivity Constraints

Having established the general existence of EF1 outcomes, in this section, we impose an additional requirement in the form of connectivity. Specifically, we assume that the goods

lie on a path $1, 2, \dots, m$. An allocation A is said to be *connected* if A_j forms an interval on the path for each $j \in [k]$. An outcome (C, A) is called *connected* if A is connected.

In the individual setting, Igarashi (2023) proved that a connected EF1 allocation is guaranteed to exist, thereby strengthening an earlier result of Bilò et al. (2022), which holds for EF2. For their proofs, Bilò et al. and Igarashi developed a discretization approach using *Sperner's lemma* (Sperner 1928). This idea was originally due to Su (1999), who provided an elegant proof that an envy-free allocation of a cake (i.e., a heterogeneous divisible good) always exists. Su's proof encodes possible divisions as points in the standard simplex and employs a triangulation of the simplex along with a coloring of each vertex of the triangulation. By applying Sperner's lemma, Su showed the existence of an elementary simplex (i.e., a simplex not composed of smaller simplices in the triangulation) whose vertices correspond to allocations that are similar to one another, where each allocation ensures that a different agent is envy-free. An infinite sequence of such simplices, with their sizes progressively shrinking, converges to an envy-free allocation.

Our main result of this section generalizes the results of Bilò et al. (2022) and Igarashi (2023) to the group setting. The result is formally stated as follows.

Theorem 2. *For any instance with arbitrary monotonic utilities such that the goods lie on a path, there exists a connected EF1 outcome.*

Note that unlike Theorem 1, this theorem does not come with efficient computation. Indeed, the question of whether a connected EF1 allocation can be computed efficiently is open even in the individual setting (Igarashi 2023).

We start by recalling some basic notions of combinatorial topology. For any positive integer k , a $(k-1)$ -simplex S is the convex hull of k main vertices $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k$. We use the notation $S = \langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$. The $(k-1)$ -standard simplex Δ^{k-1} is the $(k-1)$ -simplex whose main vertices are given by the j -th unit vectors $\mathbf{e}^j \in \{0, 1\}^k$ for $j \in [k]$, where $e_h^j = 1$ if $h = j$ and $e_h^j = 0$ otherwise.

A triangulation T of a $(k-1)$ -simplex S is a collection of smaller $(k-1)$ -simplices $S_1, S_2, \dots, S_h$ such that the union of simplices S_j for $j \in [h]$ is S , and for each $i \neq j$, the intersection $S_i \cap S_j$ is either empty or a face common to S_i and S_j . We call $S_1, S_2, \dots, S_h$ *elementary simplices*, and write $V(T)$ for the set of vertices of a triangulation T .

For a triangulation T of a $(k-1)$ -simplex S , a *coloring* is a mapping $\lambda: V(T) \rightarrow [k]$ that assigns a color in $[k]$ to each vertex $\mathbf{x} \in V(T)$. A coloring λ is called *proper* if we can write $S = \langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$ in such a way that if a vertex $\mathbf{x} \in V(T)$ is colored with color j (i.e., $j = \lambda(\mathbf{x})$), then $\mathbf{x}_j$ is a vertex of the minimal face containing $\mathbf{x}$. For example, if $k = 3$, then a $(k-1)$ -simplex can be viewed as a triangle. In a proper coloring of the simplex, the three main vertices $\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$ are colored 1, 2, 3, respectively, and each vertex $\mathbf{x}$ on an edge between $\mathbf{x}_i$ and $\mathbf{x}_j$ is colored either i or j .

In our proof, the space of connected allocations is encoded by the positions of $k-1$ “knives”. Like Bilò et al. (2022) and Igarashi (2023), we consider a triangulation of this space, where each knife is at either a vertex or an edge

of the path. More precisely, consider the following simplex:

$$S_m := \left\{ \mathbf{x} \in \mathbb{R}^{k-1} \mid \frac{1}{2} \leq x_1 \leq \dots \leq x_{k-1} \leq m + \frac{1}{2} \right\}.$$

Let T_{half} be a *Kuhn's triangulation* of S_m (Deng, Qi, and Saberi 2012; Bilò et al. 2022; Igarashi 2023). The vertices of this triangulation are given by

$$V(T_{\text{half}}) = \left\{ \mathbf{x} \in S_m \mid x_i \in \left\{ \frac{1}{2}, 1, \dots, m + \frac{1}{2} \right\}, \forall i \right\},$$

where each elementary simplex $S = \langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$ of T_{half} satisfies the property that there exists a permutation $\phi: [k] \rightarrow [k]$ such that $\mathbf{x}_{\phi(i+1)} = \mathbf{x}_{\phi(i)} + \frac{1}{2} \mathbf{e}^{\phi(i)}$ for each $i \in [k-1]$. Each vertex $\mathbf{x} \in V(T_{\text{half}})$ yields a partial allocation $A(\mathbf{x}) = (A_1(\mathbf{x}), A_2(\mathbf{x}), \dots, A_k(\mathbf{x}))$ such that for each $j \in [k]$, the j -th bundle is given by

$$A_j(\mathbf{x}) = \{y \in [m] \mid x_{j-1} < y < x_j\},$$

with $x_0 = \frac{1}{2}$ and $x_k = m + \frac{1}{2}$.

With appropriate coloring and rounding, Igarashi (2023) demonstrated that a desired elementary simplex, guaranteed by Sperner's lemma, can be rounded to produce a connected EF1 allocation. The proof relies on the notion of a *virtual utility*, $\hat{u}_a(\mathbf{x}, j)$, defined for each vertex $\mathbf{x}$ of the triangulation and each index $j \in [k]$. This virtual utility determines agent a 's most preferred bundle in a partial allocation corresponding to the vertex $\mathbf{x}$; in particular, $\hat{u}_a(\mathbf{x}, j) = 0$ if $A_j(\mathbf{x}) = \emptyset$, and $\hat{u}_a(\mathbf{x}, j) \geq 0$ otherwise. Igarashi (2023, Alg. 1) presented an algorithm that, given an elementary simplex $S = \langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$ of T_{half} , produces an allocation $A = (A_1, A_2, \dots, A_k)$ with the following property: each agent's estimate of the j -th bundle based on the virtual utility is upper-bounded by her true utility of A_j and lower-bounded by her “up-to-one utility” of A_j , as defined next. For any connected subset of goods I , the *up-to-one utility* $u_a^-(I)$ of agent a is defined as

$$u_a^-(I) := \begin{cases} 0 & \text{if } I = \emptyset; \\ \min \{u_a(I \setminus \{g\}) \mid g \in I \\ \text{such that } I \setminus \{g\} \text{ is connected} \} & \text{if } I \neq \emptyset. \end{cases}$$

The following lemma is stated as Lemma 3.2 in the work of Igarashi (2023).⁵

Lemma 1 (Igarashi 2023). *Consider the triangulation T_{half} of S_m . There exists an algorithm that, given any elementary simplex $S = \langle \mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_k \rangle$ of T_{half} , returns a connected allocation $(A_1, A_2, \dots, A_k)$ such that for every agent $a \in N$ and every pair of indices $i, j \in [k]$, we have $u_a(A_j) \geq \hat{u}_a(\mathbf{x}_i, j) \geq u_a^-(A_j)$.*

Our main lemma of this section is as follows.

Lemma 2. *Let $\lambda_1, \dots, \lambda_n: V(T_{\text{half}}) \rightarrow [k]$ be any proper colorings. Then, there exist an elementary simplex $S^* = \langle \mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_k^* \rangle$ of T_{half} and $\pi: N \rightarrow [k]$ such that*

1. $|\pi^{-1}(i)| = n_i$ for each $i \in [k]$, and

⁵Although Igarashi's Lemma 3.2 is stated for the case $k = n$, the same proof also works when $k \neq n$.

2. $\pi(a) \in \{\lambda_a(\mathbf{x}_h^*) \mid h \in [k]\}$ for each $a \in N$.

To establish Lemma 2, we will use the following result shown by Igarashi and Meunier (2025) as Lemma 4 in their work. We state a simplified version of their lemma below.⁶

Lemma 3 (Igarashi and Meunier 2025). *Consider a bipartite graph $H = (N, [k]; E)$ together with non-negative edge weights $(w_e)_{e \in E}$. Let $\delta_H(v)$ denote the set of edges incident to vertex v in H . Suppose that the following holds:*

$$\begin{aligned} \sum_{e \in \delta_H(a)} w_e &= 1 & \forall a \in N; \\ \sum_{e \in \delta_H(i)} w_e &= n_i & \forall i \in [k]. \end{aligned}$$

Then, there exists an assignment $\pi : N \rightarrow [k]$ such that

1. $|\pi^{-1}(i)| = n_i$ for each $i \in [k]$, and
2. $(a, \pi(a)) \in E$ for each $a \in N$.

Proof of Lemma 2. For each $a \in N$ and $\mathbf{x} \in V(\mathbf{T}_{\text{half}})$, we define $f^{(a)}(\mathbf{x}) = e^{\lambda_a(\mathbf{x})}$. We then extend it in an affine manner to $\bar{f}^{(a)} : S_m \rightarrow \Delta^{k-1}$ as follows. For each $\mathbf{x} \in S_m$, let $S = \langle \mathbf{x}_1, \dots, \mathbf{x}_k \rangle$ be any elementary simplex of $\mathbf{T}_{\text{half}}$ containing it. Then, write $\mathbf{x} = \sum_{h \in [k]} \alpha_h \mathbf{x}_h$, and let

$$\bar{f}^{(a)}(\mathbf{x}) = \sum_{h \in [k]} \alpha_h f^{(a)}(\mathbf{x}_h).$$

The affine extension on each simplex yields a continuous function on the entire S_m since adjacent simplices share faces and vertex values agree on those faces. Finally, we define $\bar{f} : S_m \rightarrow \Delta^{k-1}$ by letting

$$\bar{f}(\mathbf{x}) = \frac{1}{n} \sum_{a \in N} \bar{f}^{(a)}(\mathbf{x}) \quad \forall \mathbf{x} \in S_m.$$

It can be shown that $\bar{f}$ is surjective. We defer the full proof of this fact to Section A (see also, e.g., Meunier and Su 2019). Since $\bar{f}$ is surjective, there exists $\mathbf{x}^* \in S_m$ such that

$$\bar{f}(\mathbf{x}^*)_i = \frac{n_i}{n} \quad \forall i \in [k].$$

Let $S^* = \langle \mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_k^* \rangle$ be any elementary simplex of $\mathbf{T}_{\text{half}}$ containing $\mathbf{x}^*$. We define the graph $H = (N, [k]; E)$ as follows:

- $(a, i) \in E$ if and only if $\bar{f}^{(a)}(\mathbf{x}^*)_i > 0$, and
- $w_{(a,i)} = \bar{f}^{(a)}(\mathbf{x}^*)_i$.

By definition, we have

$$\sum_{e \in \delta_H(a)} w_e = \sum_{i \in [k]} \bar{f}^{(a)}(\mathbf{x}^*)_i = 1 \quad \forall a \in N$$

and

$$\sum_{e \in \delta_H(i)} w_e = \sum_{a \in N} \bar{f}^{(a)}(\mathbf{x}^*)_i = n \cdot \bar{f}(\mathbf{x}^*)_i = n_i \quad \forall i \in [k].$$

⁶We only state the case where the numbers n_i are integers, but their original lemma handles the non-integer case too.

Thus, we can apply Lemma 3 to obtain $\pi : N \rightarrow [k]$ satisfying the two conditions of the lemma. The first condition of Lemma 3 ensures the first condition of Lemma 2. Furthermore, the second condition of Lemma 3 states that $(a, \pi(a)) \in E$ for every $a \in N$. By definition of H , this implies that $\bar{f}^{(a)}(\mathbf{x}^*)_{\pi(a)} > 0$. By definition of $\bar{f}^{(a)}$, this in turn ensures that $\pi(a) = \lambda_a(\mathbf{x}_h^*)$ for some $h \in [k]$. Hence, the second condition of Lemma 2 is also satisfied. $\square$

We now show how to use Lemma 2 to prove Theorem 2.

Proof of Theorem 2. Based on the virtual utilities, define a coloring $\lambda_a : V(\mathbf{T}_{\text{half}}) \rightarrow [k]$ for each $a \in N$ such that

$$\lambda_a(\mathbf{x}) \in \operatorname{argmax}\{\hat{u}_a(\mathbf{x}, j) \mid j \in [k] \text{ such that } A_j(\mathbf{x}) \neq \emptyset\}.$$

These colorings were shown to be proper by Igarashi (2023).

Lemma 2 then yields an elementary simplex $S^* = \langle \mathbf{x}_1^*, \mathbf{x}_2^*, \dots, \mathbf{x}_k^* \rangle$ of $\mathbf{T}_{\text{half}}$ and $\pi : N \rightarrow [k]$ that satisfy the two conditions of the lemma. Let $C_i = \pi^{-1}(i)$ for each $i \in [k]$; the first condition of the lemma ensures that $|C_i| = n_i$. Next, we construct the allocation $A^* = (A_1^*, \dots, A_k^*)$ by invoking Lemma 1 on S^* . We claim that (C, A^*) is an EF1 outcome. To see this, consider any agent $a \in N$ (belonging to $C_{\pi(a)}$) and any $j \in [k]$. By the second condition of Lemma 2, $\pi(a) = \lambda_a(\mathbf{x}_h^*)$ for some $h \in [k]$. We thus have

$$\begin{aligned} u_a(A_{\pi(a)}^*) &\geq \hat{u}_a(\mathbf{x}_h^*, \pi(a)) && \text{(by Lemma 1)} \\ &\geq \hat{u}_a(\mathbf{x}_h^*, j) && \text{(since } \pi(a) = \lambda_a(\mathbf{x}_h^*)) \\ &\geq u_a^-(A_j^*) && \text{(by Lemma 1).} \end{aligned}$$

Hence, (C, A^*) is an EF1 outcome, as desired. $\square$

5 Envy-Freeness: Asymptotic Existence

In this section, we turn our attention to the asymptotic existence of envy-free outcomes. Specifically, we consider a setting where the utilities are additive and, for each $a \in N$ and $g \in M$, the utility $u_a(g)$ is drawn independently from a given distribution $\mathcal{D}$. This distribution is assumed to be non-atomic (i.e., does not put a positive probability on any single point) and has a support contained in the interval⁷ $[0, 1]$. We also assume throughout this section that all groups have the same size, that is, $n_1 = \dots = n_k = n/k$ —in particular, n is divisible by k . We fix k and investigate the asymptotic (non-)existence of EF outcomes as n grows. We say that an event occurs *with high probability* if the probability that it occurs approaches 1 as $n \rightarrow \infty$.

5.1 Divisible Case

We first consider the case where the number of goods m is divisible by k . Intuitively, this is an “easier” case, as it is possible to give every group the same number of goods. For this case, we show that the phase transition occurs at $m = \Theta(\log n)$, as stated in the following two theorems.

Theorem 3. *For any fixed $k \geq 2$, if $m \leq \frac{\ln n}{4}$, then with high probability, no envy-free outcome exists.*

⁷If the support is contained in $[0, t]$ for some $t > 1$, we can scale down all utilities by t .

Theorem 4. For any fixed $k \geq 2$, utility distribution $\mathcal{D}$, and $\beta \in (0, 1)$, there exists a constant $C_{k,\beta,\mathcal{D}}$ such that, for any sufficiently large n and any $m \geq C_{k,\beta,\mathcal{D}} \cdot \ln n$ with the property that m is divisible by k , there is a polynomial-time algorithm that computes an envy-free outcome with probability at least $1 - \beta$.

To facilitate the proofs, we introduce some definitions.

- Let $\mathcal{A}$ denote the set of all allocations.
- An allocation $A = (A_1, \dots, A_k) \in \mathcal{A}$ is called *balanced* if $|A_1| = \dots = |A_k| = m/k$. Let $\mathcal{A}_{\text{bal}}$ be the set of all balanced allocations.
- For $A \in \mathcal{A}$ and $a \in N$, let $i_a^*(A) = \operatorname{argmax}_{i \in [k]} u_a(A_i)$, ties broken arbitrarily. For each $i \in [k]$, let $p_i^A = \Pr[i_a^*(A) = i]$, where the probability is taken over the randomness of the utilities. Let $\mathbf{p}^A = (p_1^A, \dots, p_k^A)$.
- Let $\mathbf{C}^{\text{EF},A}$ be the partition of N where each agent $a \in N$ is assigned to $i_a^*(A)$. Note that the outcome $(\mathbf{C}^{\text{EF},A}, A)$ is always envy-free.
- Let $\mathbb{P}(N)$ denote the set of all partitions $(C_1, \dots, C_k)$ of N such that $|C_1| = \dots = |C_k| = n/k$.
- For $A \in \mathcal{A}$, let $\mathcal{E}_A$ denote the event that $\mathbf{C}^{\text{EF},A} \in \mathbb{P}(N)$.

Since $\mathcal{D}$ is non-atomic, a tie in utilities (i.e., $u_a(B) = u_a(B')$ for some agent a and distinct bundles B, B') occurs with probability zero. We thus have the following.

Observation 1. The probability that an envy-free outcome exists is equal to $\Pr[\bigvee_{A \in \mathcal{A}} \mathcal{E}_A]$.

We continue with further preliminaries on probabilities.

- For a distribution $\mathcal{P}$, we write $\operatorname{supp}(\mathcal{P})$ to denote its support. For $\Lambda \subseteq \operatorname{supp}(\mathcal{P})$, we write $\mathcal{P}(\Lambda)$ as a shorthand for the measure (i.e., probability) of Λ under $\mathcal{P}$. When $\Lambda = \{\lambda\}$ has size one, we simply write $\mathcal{P}(\lambda)$.
- For $\Sigma \in \mathbb{R}^{d \times d}$, we write $\mathcal{N}(\mathbf{0}, \Sigma)$ to denote the centered (multivariate) Gaussian distribution supported on $\mathbb{R}^d$ with covariance Σ .
- Let Δ_n^{k-1} denote the discrete $(k-1)$ -simplex $\{\mathbf{x} \in \mathbb{Z}_{\geq 0}^k \mid \sum_{i \in [k]} x_i = n\}$.
- For $n \in \mathbb{N}$ and $\mathbf{p} \in \Delta^{k-1}$, the multinomial distribution $\operatorname{Mult}(n, \mathbf{p})$ is a distribution supported on Δ_n^{k-1} where $\operatorname{Mult}(n, \mathbf{p})\{\mathbf{x}\} := \binom{n}{\mathbf{x}} \prod_{i \in [k]} p_i^{x_i}$ for each $\mathbf{x} \in \Delta_n^{k-1}$.

We do not include every proof here; all missing proofs can be found in the appendix.

Non-Existence via First Moment Method. We begin by proving the non-existence result (Theorem 3) via the first moment method. Specifically, in light of Observation 1, we compute $\Pr[\mathcal{E}_A]$ for any allocation A and apply the union bound. We start with a simple formula.

Observation 2. For any allocation A , we have $\Pr[\mathcal{E}_A] = \operatorname{Mult}(n, \mathbf{p}^A) \left\{ \frac{n}{k} \cdot \mathbf{1} \right\}$.

Proof. Since the agents' utilities are independent, the distribution of the number of agents in $\mathbf{C}^{\text{EF},A}$ is $\operatorname{Mult}(n, \mathbf{p}^A)$. The formula follows from the definition of $\mathcal{E}_A$. $\square$

A standard bound on $\operatorname{Mult}(n, \mathbf{p}^A) \left\{ \frac{n}{k} \cdot \mathbf{1} \right\}$ then yields the following lemma.

Lemma 4. For any $A \in \mathcal{A}$, we have $\Pr[\mathcal{E}_A] \leq \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}}$. Moreover, for any $A \in \mathcal{A}_{\text{bal}}$, $\Pr[\mathcal{E}_A] \geq \frac{1}{e^{k^2/n}} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}}$.

We can now prove Theorem 3 by taking the union bound.

Proof of Theorem 3. Using Observation 1 and Lemma 4 and taking the union bound over all $A \in \mathcal{A}$, we find that the probability that an envy-free outcome exists is at most

$$k^m \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}} \leq n^{\frac{\ln k}{4}} \cdot \frac{k^{k/2}}{n^{\frac{k-1}{2}}} = \frac{k^{k/2}}{n^{\frac{2k - \ln k - 2}{4}}},$$

where we use the assumption that $m \leq \frac{\ln n}{4}$. The last quantity is $o(1)$ for any fixed $k \geq 2$. $\square$

Existence via Second Moment Method. Next, we turn to the existence result (Theorem 4). Recall that we have estimated the first moment $\Pr[\mathcal{E}_A]$ in Lemma 4. One might expect that when the sum of $\Pr[\mathcal{E}_A]$ over all $A \in \mathcal{A}$ far exceeds 1, our desired event $\bigvee_{A \in \mathcal{A}} \mathcal{E}_A$ occurs with high probability. However, this is not necessarily true for dependent events. To overcome this issue, we employ the *second moment method*, which leverages the fact that if the events are nearly independent, in the sense that $\Pr[\mathcal{E}_A \wedge \mathcal{E}_{A'}] \leq (1 + o(1)) \Pr[\mathcal{E}_A] \cdot \Pr[\mathcal{E}_{A'}]$, then we can reach the desired conclusion. Unfortunately, such an inequality need not hold for all pairs of allocations A, A' . Indeed, when A and A' are “close”, $\mathcal{E}_A$ and $\mathcal{E}_{A'}$ can be highly correlated. Therefore, we will only apply the second-moment method on a set of allocations that are “far away” from one another. The exact definition of “far away” that we use is given below in Definition 1. Observe that if A, A' are drawn independently at random from $\mathcal{A}_{\text{bal}}$, then $\mathbb{E}[|A_i \cap A'_j|] = m/k^2$ for any $i, j \in [k]$. Thus, the condition in Definition 1 requires that $|A_i \cap A'_j|$ is close to its expectation.

Definition 1. For $\delta \in [0, 1]$, two balanced allocations $A = (A_1, \dots, A_k)$ and $A' = (A'_1, \dots, A'_k)$ are said to be δ -intersection balanced (δ -IB) if $\frac{m(1-\delta)}{k^2} \leq |A_i \cap A'_j| \leq \frac{m(1+\delta)}{k^2}$ for all $i, j \in [k]$. We say that balanced allocations $A^{(1)}, \dots, A^{(D)}$ are δ -IB if every pair among them is δ -IB.

A standard concentration argument shows that if $m \geq \Theta_{k,\delta}(\log n)$, then n^{2k} random balanced allocations satisfy δ -IB with high probability, as stated below.

Lemma 5. For any $\delta \in (0, 1)$ and any $m \geq \frac{4k^4}{\delta^2} \cdot \ln n$ such that m is divisible by k , let $A^{(1)}, \dots, A^{(n^{2k})}$ be balanced allocations drawn uniformly and independently at random. Then, with high probability, these allocations are δ -IB.

The remainder of this section is largely devoted to bounding the second moment as stated in the following lemma, where $\sigma^2 > 0$ is the variance of $\mathcal{D}$.

Lemma 6. For any fixed $\beta \in (0, 1)$, if $\delta \leq \frac{\beta}{12k^4}$ and $m \geq \frac{48000k^{13}}{\beta^2\sigma^6}$, then for any pair of δ -IB balanced allocations A

and A' , we have

$$\Pr[\mathcal{E}_{A'} \mid \mathcal{E}_A] \leq \frac{1}{1 - \beta/2 - o(1)} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}}.$$

Before we prove Lemma 6, let us first show how to use it to establish Theorem 4.

Proof of Theorem 4. Let $C_{k,\beta,\mathcal{D}} = \max \left\{ \frac{4k^4}{\delta^2}, \frac{1920000k^{13}}{\beta^2\sigma^6} \right\}$, where $\delta = \frac{\beta}{12k^4}$. The algorithm works as follows.

- Choose balanced allocations $A^{(1)}, \dots, A^{(n^{2k})}$ independently and uniformly at random.⁸
- For $w \in [n^{2k}]$, check whether $\mathbf{C}^{\text{EF},A^{(w)}} \in \mathbb{P}(N)$. If so, output $(\mathbf{C}^{\text{EF},A^{(w)}}, A^{(w)})$.

The algorithm runs in time $n^{O(k)} \cdot m$. To bound its success probability, note that by Lemma 5, $A^{(1)}, \dots, A^{(n^{2k})}$ are δ -IB with high probability. We condition on this event henceforth.

Our algorithm succeeds when the event $\bigvee_{w \in [n^{2k}]} \mathcal{E}_{A^{(w)}}$ occurs. Let $Z = |\{w \in [n^{2k}] \mid \mathcal{E}_{A^{(w)}}\}|$ and $q = \Pr[\mathcal{E}_{A^{(w)}}]$; note that q does not depend on w due to symmetry. Using the second moment method, the probability that the algorithm succeeds can be bounded as follows:

$$\begin{aligned} \Pr \left[\bigvee_{w \in [n^{2k}]} \mathcal{E}_{A^{(w)}} \right] &= \Pr[Z > 0] \geq \frac{(\mathbb{E}[Z])^2}{\mathbb{E}[Z^2]} \\ &= \frac{1}{\frac{1}{n^{4k}} \sum_{w \in [n^{2k}]} \left(\frac{1}{q} + \sum_{v \in [n^{2k}] \setminus \{w\}} \frac{\Pr[\mathcal{E}_{A^{(v)}} \mid \mathcal{E}_{A^{(w)}}]}{q} \right)}. \end{aligned}$$

Applying Lemmas 4 and 6, the denominator is at most

$$\begin{aligned} \frac{1}{n^{4k}} \sum_{w \in [n^{2k}]} \left(o \left(n^{\frac{k-1}{2}} \right) + \sum_{v \in [n^{2k}] \setminus \{w\}} \frac{1}{1 - \beta/2 - o(1)} \right) \\ = \frac{1}{1 - \beta/2 - o(1)}. \end{aligned}$$

Thus, our algorithm succeeds with probability $1 - \beta/2 - o(1)$, which is at least $1 - \beta$ for any sufficiently large n . $\square$

We now proceed to prove Lemma 6. Fix two balanced allocations A, A' that are δ -IB. For $i, i' \in [k]$, we let⁹ $p_{(i,i')} = \Pr[i_a^*(A) = i \wedge i_a^*(A') = i']$.

Lemma 7. *For each $a \in N$, let X^a be a random variable on $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ such that $\Pr[X^a = \mathbf{e}_{i'}] = k \cdot p_{(\lceil ka/n \rceil, i')}$ for each $i' \in [k]$. Then, $\Pr[\mathcal{E}_{A'} \mid \mathcal{E}_A] = \Pr[X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}]$.*

Proof. By definition, $\Pr[\mathcal{E}_{A'} \mid \mathcal{E}_A] = \Pr[\mathbf{C}^{\text{EF},A'} \in \mathbb{P}(N) \mid \mathbf{C}^{\text{EF},A} \in \mathbb{P}(N)]$. Due to the symmetry across agents and since A is balanced, we may assume that $\mathbf{C}^{\text{EF},A}$ assigns agent $a \in N$ to group $\lceil ka/n \rceil$. Conditioned on this, we can

⁸For example, we can take a random permutation of the goods, and let each block of m/k goods form a bundle.

⁹ $p_{(i,i')}$ is the same for every agent a , so we omit a from the notation.

take X^a to be the random variable such that $X^a = \mathbf{e}_{i_a^*(A')}$. Indeed, we have that

$$\begin{aligned} \Pr[X^a = \mathbf{e}_{i'}] &= \Pr[i_a^*(A') = i' \mid i_a^*(A) = \lceil ka/n \rceil] \\ &= \frac{p_{(\lceil ka/n \rceil, i')}}{\Pr[i_a^*(A) = \lceil ka/n \rceil]} = k \cdot p_{(\lceil ka/n \rceil, i')}, \end{aligned}$$

where we use Bayes' law and the symmetry across groups, respectively. Finally, observe that $\mathbf{C}^{\text{EF},A'} \in \mathbb{P}(N)$ is equivalent to $X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}$. $\square$

The rest of the proof of Lemma 6 can be divided into two parts: (i) showing that the values $p_{(i,i')}$ are all close to $1/k^2$, and (ii) bounding $\Pr[X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}]$.

Part (i): Bounding $p_{(i,i')}$. We need the following multivariate generalization of the Berry–Esseen theorem, which is stated as Theorem 1.1 in the work of Raič (2019).¹⁰

Lemma 8 (Raič 2019). *Let $W^1, \dots, W^T$ be independent random variables in $\mathbb{R}^d$ with mean zero, $H := W^1 + \dots + W^T$, and $\Sigma := \text{Cov}[H] \in \mathbb{R}^{d \times d}$ be the covariance matrix of H . For any convex $\Lambda \subseteq \mathbb{R}^d$, it holds that $|\Pr[H \in \Lambda] - \mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda\}| \leq 60d^{1/4} \sum_{i \in [T]} \mathbb{E}[\|\Sigma^{-\frac{1}{2}} W^i\|_2^3]$.*

Let us now give the proof overview for this part. We let $d = k^2$ and implicitly associate¹¹ tuples $(j, j') \in [k]^2$ with elements in $[k^2]$ when writing the indices for readability.

Let μ denote the mean of $\mathcal{D}$. For each $g \in M$, we define the random variable $W^g \in \mathbb{R}^d$ such that

$$W_{(j,j')}^g = \begin{cases} (u_a(g) - \mu) \cdot \frac{k}{\sigma\sqrt{m}} & \text{if } g \in A_j \cap A'_{j'}; \\ 0 & \text{otherwise} \end{cases}$$

for all $j, j' \in [k]$; note that this distribution is the same regardless of a . Then, let $H = \sum_{g \in M} W^g$. Moreover, for all $i, i' \in [k]$, let $\Lambda_{i,i'} \subseteq \mathbb{R}^d$ be the set of all vectors $\mathbf{v} \in \mathbb{R}^d$ that satisfy the following constraints:

$$\begin{aligned} \sum_{\ell \in [k]} v_{(i,\ell)} &\geq \sum_{\ell \in [k]} v_{(j,\ell)} & \forall j \in [k]; \\ \sum_{\ell \in [k]} v_{(\ell,i')} &\geq \sum_{\ell \in [k]} v_{(\ell,j')} & \forall j' \in [k]. \end{aligned}$$

One can check that when there are no ties in utilities, $i_a^*(A) = i$ and $i_a^*(A') = i'$ if and only if $H \in \Lambda_{i,i'}$. In other words, $\Pr[H \in \Lambda_{i,i'}]$ is exactly $p_{(i,i')}$. Since $\Lambda_{i,i'}$ is convex, Lemma 8 implies that $p_{(i,i')}$ is close to $\mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\}$ for $\Sigma = \text{Cov}[H]$. Using the fact that A, A' are δ -IB and balanced, we can show that Σ is also close to the identity matrix I_d . By applying a standard total variation distance bound between Gaussians, this also implies that $\mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\}$ and $\mathcal{N}(\mathbf{0}, I_d)\{\Lambda_{i,i'}\}$ are close. Due to symmetry, the latter is simply $\frac{1}{k^2}$. Thus, we can conclude that $p_{(i,i')}$ itself is close to $\frac{1}{k^2}$, as stated more formally below.

¹⁰The version stated here follows from Theorem 1.1 of Raič (2019) by letting $X_i = \Sigma^{-\frac{1}{2}} W^i$, so that $\sum_{i=1}^T \text{Var}(X_i) = I_d$. Also, we use the term $60d^{1/4}$, which is weaker than the one used by Raič (2019).

¹¹For example, (j, j') can be associated with $k(j-1) + j'$.

Lemma 9. Let $\gamma \in (0, 1)$. Suppose that $\delta \leq \frac{\gamma}{3k}$ and $m \geq \frac{30000k^7}{\gamma^2 \sigma^6}$. Then, for all $i, i' \in [k]$, we have $|p_{(i,i')} - \frac{1}{k^2}| \leq \gamma$.

Part (ii): Bounding $\Pr[X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}]$. We reinterpret X^i so that it is a mixture distribution between the uniform distribution on $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ and a “leftover” distribution. In other words, for each X^i , we can toss a (biased) coin and, based on the outcome, sample X^i either from the uniform distribution or from the leftover distribution. By a concentration bound, we show that X^i is drawn from the uniform distribution for most indices i . For these X^i , their sum exactly follows the multinomial distribution, which we have a very good estimate on. From this, we can derive the following bound.

Lemma 10. Let $\zeta \in (0, 0.5)$ and $\tilde{X}^1, \dots, \tilde{X}^n$ be any independent random variables on $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ such that $|\Pr[\tilde{X}^a = \mathbf{e}_i] - \frac{1}{k}| \leq \frac{\zeta}{k}$ for all $a \in N$ and $i \in [k]$. If $n \geq \frac{k}{1-2\zeta}$, then $\Pr[\tilde{X}^1 + \dots + \tilde{X}^n = \frac{n}{k} \cdot \mathbf{1}]$ is at most

$$\exp(-2\zeta^2 n) + \frac{k^{k/2}}{(2\pi(1-2\zeta)n)^{\frac{k-1}{2}}}.$$

Finally, combining Lemmas 7, 9, and 10 yields Lemma 6.

5.2 Non-Divisible Case

Next, we consider the case where m is not divisible by k . For this case, we prove that an envy-free outcome cannot exist with high probability unless $m = \Omega(\sqrt{n})$. This differs markedly from the divisible case, where $m = \omega(\log n)$ suffices for existence.

Theorem 5. For any fixed $k \geq 2$, if $m = o(\sqrt{n})$ and m is not divisible by k , then with high probability, no envy-free outcome exists.

The proof of this non-existence result, like the divisible case (Theorem 3), uses the first moment method. The key distinguishing property between the two cases is encapsulated in the lemma below, which states that for any allocation A , there exists $i \in [k]$ such that p_i^A is noticeably smaller than $\frac{1}{k}$; in particular, we may take an index i corresponding to a bundle with the smallest size. This is in contrast to the divisible case, for which a balanced allocation gives $p_i^A = \frac{1}{k}$ for all $i \in [k]$.

Lemma 11. If m is not divisible by k , then for any $A \in \mathcal{A}$, there exists $i \in [k]$ such that $p_i^A \leq \frac{1}{k} - \frac{\alpha}{k^2 \sqrt{m}}$, where $\alpha > 0$ is a constant depending only on $\mathcal{D}$.

Lemma 11 implies the following upper bound on $\Pr[\mathcal{E}_A]$.

Lemma 12. If m is not divisible by k , then for any $A \in \mathcal{A}$, we have $\Pr[\mathcal{E}_A] \leq \sqrt{k} \cdot \exp\left(-\frac{\alpha^2}{k^4} \cdot \frac{n}{m}\right)$, where α is the constant from Lemma 11.

We now finish the proof in a similar manner as Theorem 3.

Proof of Theorem 5. Using Observation 1 and Lemma 12 and taking the union bound over all $A \in \mathcal{A}$, the probability that an envy-free outcome exists is at most $k^m \cdot \sqrt{k} \cdot \exp\left(-\frac{\alpha^2}{k^4} \cdot \frac{n}{m}\right)$, which is $o(1)$ because $m = o(n/m)$. $\square$

When $m = \Omega(n \log n)$, an envy-free outcome exists with high probability due to the result in the fixed-group setting by Manurangsi and Suksompong (2017). Tightening the gap between $o(\sqrt{n})$ and $\Omega(n \log n)$ is an interesting question.

6 Conclusion

In this paper, we have studied fairness in the allocation of indivisible goods with variable groups, where a partition of the agents into groups can be chosen along with an allocation of the goods. We demonstrated that the flexibility afforded by this model allows strong envy-freeness guarantees to be made in both worst-case and average-case scenarios.

Besides closing the asymptotic gap in the non-divisible case (Section 5.2) and extending the asymptotic analysis in Section 5 to the case where groups may have differing sizes, an intriguing direction for future work is to make connections between the variable-group model and the setting of *hedonic games* (Aziz and Savani 2016). Specifically, in hedonic games, agents derive utilities from other agents assigned to the same group, and the objective is to find a desirable partition of the agents into groups; variants where the group sizes are fixed have also been considered (Bild, Monaco, and Moscardelli 2022; Li et al. 2023). A generalization of both models would therefore be to permit preferences both over agents as well as over goods—this can reflect, e.g., group projects where resources are assigned to each group.¹² It would be interesting to investigate fairness guarantees that can be made in this general setting.

Acknowledgments

This work was partially supported by the Singapore Ministry of Education under grant number MOE-T2EP20221-0001, by JST FOREST Grant Number JPMJFR226O, and by an NUS Start-up Grant. We would like to thank Erel Segal-Halevi for helpful discussions at the early stage, and the anonymous reviewers for their valuable comments.

References

- Artin, E. 1964. *The Gamma Function*. Athena Series. Holt, Rinehart and Winston.
- Aziz, H. 2020. Developments in multi-agent fair allocation. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*, 13563–13568.
- Aziz, H.; and Savani, R. 2016. Hedonic games. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*, chapter 15, 356–376. Cambridge University Press.
- Bai, Y.; Feige, U.; Gözl, P.; and Procaccia, A. D. 2022. Fair allocations for smoothed utilities. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC)*, 436–465.
- Bai, Y.; and Gözl, P. 2022. Envy-free and Pareto-optimal allocations for agents with asymmetric random valuations.

¹²A similar model has been studied under the name *generalized group activity selection problem* (Bild et al. 2019; Flammini and Varricchio 2022).

- In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 53–59.
- Barman, S.; Ebadian, S.; Latifian, M.; and Shah, N. 2025. Fair division with market values. In *Proceedings of the 39th AAAI Conference on Artificial Intelligence (AAAI)*, 13589–13596.
- Bei, X.; Igarashi, A.; Lu, X.; and Suksompong, W. 2022. The price of connectivity in fair division. *SIAM Journal on Discrete Mathematics*, 36(2): 1156–1186.
- Benadè, G.; Halpern, D.; Psomas, A.; and Verma, P. 2024. On the existence of envy-free allocations beyond additive valuations. In *Proceedings of the 25th ACM Conference on Economics and Computation (EC)*, 1287.
- Bilò, V.; Caragiannis, I.; Flammini, M.; Igarashi, A.; Monaco, G.; Peters, D.; Vinci, C.; and Zwicker, W. S. 2022. Almost envy-free allocations with connected bundles. *Games and Economic Behavior*, 131: 197–221.
- Bilò, V.; Fanelli, A.; Flammini, M.; Monaco, G.; and Moscardelli, L. 2019. Optimality and Nash stability in additive separable generalized group activity selection problems. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, 102–108.
- Bilò, V.; Monaco, G.; and Moscardelli, L. 2022. Hedonic games with fixed-size coalitions. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*, 9287–9295.
- Bouveret, S.; Cechlárová, K.; Elkind, E.; Igarashi, A.; and Peters, D. 2017. Fair division of a graph. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, 135–141.
- Bouveret, S.; Chevaleyre, Y.; and Maudet, N. 2016. Fair allocation of indivisible goods. In Brandt, F.; Conitzer, V.; Endriss, U.; Lang, J.; and Procaccia, A. D., eds., *Handbook of Computational Social Choice*, chapter 12, 284–310. Cambridge University Press.
- Bu, X.; Li, Z.; Liu, S.; Song, J.; and Tao, B. 2023. Fair division with allocator’s preference. In *Proceedings of the 19th International Conference on Web and Internet Economics (WINE)*, 77–94.
- Caragiannis, I.; Larsen, K. G.; and Shyam, S. 2025. A new lower bound for multi-color discrepancy with applications to fair division. In *Proceedings of the 18th International Symposium on Algorithmic Game Theory (SAGT)*, 228–246.
- Deng, X.; Qi, Q.; and Saberi, A. 2012. Algorithmic solutions for envy-free cake cutting. *Operations Research*, 60(6): 1461–1476.
- Devroye, L.; Mehrabian, A.; and Reddad, T. 2023. The total variation distance between high-dimensional Gaussians with the same mean. *CoRR*, abs/1810.08693v7.
- Dickerson, J. P.; Goldman, J.; Karp, J.; Procaccia, A. D.; and Sandholm, T. 2014. The computational rise and fall of fairness. In *Proceedings of the 28th AAAI Conference on Artificial Intelligence (AAAI)*, 1405–1411.
- Flammini, M.; and Varricchio, G. 2022. Approximate strategyproof mechanisms for the additively separable group activity selection problem. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, 300–306.
- Ghods, M.; Latifian, M.; Mohammadi, A.; Moradian, S.; and Seddighin, M. 2018. Rent division among groups. In *Proceedings of the 12th International Conference on Combinatorial Optimization and Applications (COCOA)*, 577–591.
- Gölz, P.; and Yaghoubizade, H. 2026. Fair division among couples and small groups. In *Proceedings of the 40th AAAI Conference on Artificial Intelligence (AAAI)*. Forthcoming.
- Igarashi, A. 2023. How to cut a discrete cake fairly. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 5681–5688.
- Igarashi, A.; and Meunier, F. 2025. Envy-free division of multilayered cakes. *Mathematics of Operations Research*, 50(3): 2261–2286.
- Kawase, Y.; Roy, B.; and Sanpui, M. A. 2025. Simultaneously fair allocation of indivisible items across multiple dimensions. In *Proceedings of the 45th IARCS Annual Conference on Foundations of Software Technology and Theoretical Computer Science (FSTTCS)*. Forthcoming.
- Kurokawa, D.; Procaccia, A. D.; and Wang, J. 2016. When can the maximin share guarantee be guaranteed? In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, 523–529.
- Kyropoulou, M.; Suksompong, W.; and Voudouris, A. A. 2020. Almost envy-freeness in group resource allocation. *Theoretical Computer Science*, 841: 110–123.
- Li, L.; Micha, E.; Nikolov, A.; and Shah, N. 2023. Partitioning friends fairly. In *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI)*, 5747–5754.
- Lipton, R. J.; Markakis, E.; Mossel, E.; and Saberi, A. 2004. On approximately fair allocations of indivisible goods. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC)*, 125–131.
- Manurangsi, P.; and Meka, R. 2026. Tight lower bound for multicolor discrepancy. In *Proceedings of the 9th SIAM Symposium on Simplicity in Algorithms (SOSA)*. Forthcoming.
- Manurangsi, P.; and Suksompong, W. 2017. Asymptotic existence of fair divisions for groups. *Mathematical Social Sciences*, 89: 100–108.
- Manurangsi, P.; and Suksompong, W. 2020. When do envy-free allocations exist? *SIAM Journal on Discrete Mathematics*, 34(3): 1505–1521.
- Manurangsi, P.; and Suksompong, W. 2021. Closing gaps in asymptotic fair division. *SIAM Journal on Discrete Mathematics*, 35(2): 668–706.
- Manurangsi, P.; and Suksompong, W. 2022. Almost envy-freeness for groups: Improved bounds via discrepancy theory. *Theoretical Computer Science*, 930: 179–195.
- Manurangsi, P.; and Suksompong, W. 2025a. Asymptotic fair division: Chores are easier than goods. In *Proceedings of the 34th International Joint Conference on Artificial Intelligence (IJCAI)*, 3988–3995.

Manurangsi, P.; and Suksompong, W. 2025b. Ordinal maximin guarantees for group fair division. *Theoretical Computer Science*, 1036: 115151.

Manurangsi, P.; Suksompong, W.; and Yokoyama, T. 2025. Asymptotic analysis of weighted fair division. *Theoretical Computer Science*, 1054: 115533.

Markakis, E. 2017. Approximation algorithms and hardness results for fair division. In Endriss, U., ed., *Trends in Computational Social Choice*, chapter 12, 231–247. AI Access.

Meunier, F.; and Su, F. E. 2019. Multilabeled versions of Sperner’s and Fan’s lemmas and applications. *SIAM Journal on Applied Algebra and Geometry*, 3(3): 391–411.

Postnikov, A. G.; and Yudin, A. A. 1988. Estimating the maximum probability of a sum of independent vectors. *Theory of Probability & Its Applications*, 32(2): 331–334.

Raič, M. 2019. A multivariate Berry–Esseen theorem with explicit constants. *Bernoulli*, 25(4A): 2824–2853.

Robbins, H. 1955. A remark on Stirling’s formula. *American Mathematical Monthly*, 62(1): 26–29.

Segal-Halevi, E.; and Nitzan, S. 2019. Envy-free cake-cutting among families. *Social Choice and Welfare*, 53(4): 709–740.

Segal-Halevi, E.; and Suksompong, W. 2021. How to cut a cake fairly: a generalization to groups. *American Mathematical Monthly*, 128(1): 79–83.

Segal-Halevi, E.; and Suksompong, W. 2023. Cutting a cake fairly for groups revisited. *American Mathematical Monthly*, 130(3): 203–213.

Serfling, R. J. 1974. Probability inequalities for the sum in sampling without replacement. *The Annals of Statistics*, 2(1): 39–48.

Sperner, E. 1928. Neuer Beweis für die Invarianz der Dimensionszahl und des Gebietes. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*, 6: 265–272.

Su, F. E. 1999. Rental harmony: Sperner’s lemma in fair division. *American Mathematical Monthly*, 106(10): 930–942.

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

Ushakov, N. G. 1986. Upper estimates of maximum probability for sums of independent random vectors. *Theory of Probability & Its Applications*, 30(1): 38–49.

Walsh, T. 2020. Fair division: the computer scientist’s perspective. In *Proceedings of the 29th International Joint Conference on Artificial Intelligence (IJCAI)*, 4966–4972.

Yokoyama, T.; and Igarashi, A. 2025. Asymptotic existence of class envy-free matchings. In *Proceedings of the 24th International Conference on Autonomous Agents and Multi-agent Systems (AAMAS)*, 2244–2252.

A Surjectivity of $\bar{f}$ in the proof of Lemma 2

We now show the surjectivity of $\bar{f}$ as defined in the proof of Lemma 2. To this end, we will use the following lemma, which is stated as Lemma 2.8 in the work of Meunier and Su (2019).

Lemma 13 (Meunier and Su 2019). *Any continuous map q from a polytope to itself, which satisfies $q(F) \subseteq F$ for every face F of the polytope, is surjective.*

Lemma 14. *Let $\bar{f}$ be as defined in the proof of Lemma 2. Then, $\bar{f}$ is surjective.*

Proof. Let $\psi : \Delta^{k-1} \rightarrow S_m$ denote the canonical mapping from Δ^{k-1} to S_m , i.e.,

$$\psi(\mathbf{y})_i = \frac{1}{2} + m \cdot \sum_{j=1}^i y_j \quad \forall i \in [k-1].$$

Thus, to show that $\bar{f}$ is surjective, it suffices to show that $q := \bar{f} \circ \psi$ is surjective. By Lemma 13, we only need to show that q satisfies the condition of the lemma. For each $a \in N$, define $q^{(a)} := \bar{f}^{(a)} \circ \psi$. Since $q = \frac{1}{n} \sum_{a \in N} q^{(a)}$, it is in turn sufficient to show that each $q^{(a)}$ satisfies the condition of Lemma 13.

Let us now fix $a \in N$, a face F of Δ^{k-1} , and $\mathbf{y} \in F$. Consider $\mathbf{x} = \psi(\mathbf{y})$. By definition of $\bar{f}^{(a)}$, there exist¹³ $\mathbf{x}^1, \dots, \mathbf{x}^{k'} \in V(T_{\text{half}})$ and $\alpha^1, \dots, \alpha^{k'} > 0$ such that

$$\begin{aligned} \sum_{h \in [k']} \alpha^h &= 1; \\ \sum_{h \in [k']} \alpha^h \mathbf{x}^h &= \mathbf{x}; \\ \sum_{h \in [k']} \alpha^h f^{(a)}(\mathbf{x}^h) &= \bar{f}^{(a)}(\mathbf{x}). \end{aligned}$$

Thus, to show that $q^{(a)}(\mathbf{y}) = \bar{f}^{(a)}(\mathbf{x})$ belongs to F , it suffices to argue that $f^{(a)}(\mathbf{x}^h)$ belongs to F for all $h \in [k']$.

Since F is a face of Δ^{k-1} , there exists $S \subseteq [k]$ such that F is the span of $\{\mathbf{e}^j \mid j \in S\}$. From the definition of ψ , this implies that $x_i = x_{i-1}$ for all $i \notin S$. Since $\mathbf{x}$ is a strictly positive convex combination of $\mathbf{x}^1, \dots, \mathbf{x}^h$, we also have that $x_i^h = x_{i-1}^h$ for all $h \in [k']$ and $i \notin S$. This means that $\mathbf{x}^h$ belongs to the span of $\{\psi(\mathbf{e}^j) \mid j \in S\}$; note also that each $\psi(\mathbf{e}^j)$ is a vertex of $V(T_{\text{half}})$. Since λ_a is proper, we have that $\lambda_a(\mathbf{x}^h) \in S$. It follows that $f^{(a)}(\mathbf{x}^h) = \mathbf{e}^{\lambda_a(\mathbf{x}^h)}$ belongs to F , as desired. $\square$

B Deferred Proofs from Section 5

B.1 Additional Preliminaries

We introduce some additional notation.

¹³Note that k' can be strictly less than k if $\mathbf{x}$ belong to a face of an elementary simplex (since we only consider those $\mathbf{x}^h$ with coefficient strictly greater than zero).

- For $p \in [0, 1]$ and $r \in \mathbb{N}$, let $\text{Ber}(p)$ and $\text{Bin}(r, p)$ denote the Bernoulli and binomial distributions with success probability p , respectively.
- We use D_{KL} and D_{TV} to denote the KL divergence and the total variation (TV) distance of two distributions, respectively. Recall that the TV distance of two distributions $\mathcal{P}_1, \mathcal{P}_2$ is defined by

$$D_{\text{TV}}(\mathcal{P}_1, \mathcal{P}_2) = \sup_{\Lambda \subseteq \text{supp}(\mathcal{P}_1) \cup \text{supp}(\mathcal{P}_2)} |\mathcal{P}_1(\Lambda) - \mathcal{P}_2(\Lambda)|.$$

Since we will only use the KL divergence on discrete distributions, we only state the definition for such distributions:

$$D_{\text{KL}}(\mathcal{P}_1 \| \mathcal{P}_2) = \sum_{\lambda \in \text{supp}(\mathcal{P}_1)} \mathcal{P}_1(\lambda) \cdot \ln \left(\frac{\mathcal{P}_1(\lambda)}{\mathcal{P}_2(\lambda)} \right).$$

We will also use Pinsker's inequality, which relates the TV distance and the KL divergence.

Lemma 15 (Pinsker's inequality). *For any distributions $\mathcal{P}_1$ and $\mathcal{P}_2$, it holds that $D_{\text{TV}}(\mathcal{P}_1, \mathcal{P}_2) \leq \sqrt{\frac{1}{2} \cdot D_{\text{KL}}(\mathcal{P}_1 \| \mathcal{P}_2)}$.*

B.2 Multinomial Probability Mass Estimate

The following estimate of the probability mass of the multinomial distribution follows from Stirling's approximation of factorials. It will be useful in the proof of both the existential and non-existential results.

Lemma 16. *For any $n, k \in \mathbb{N}$ such that n is divisible by k and any $\mathbf{p} \in \Delta^{k-1}$, there exists $\tilde{r}_{n,k} \in [0, \frac{k^2}{n}]$ such that*

$$\text{Mult}(n, \mathbf{p}) \left\{ \frac{n}{k} \cdot \mathbf{1} \right\} = \frac{1}{e^{\tilde{r}_{n,k}}} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}} \cdot \frac{1}{e^{n \cdot D_{\text{KL}}(\frac{1}{k} \| \mathbf{p})}}.$$

To prove Lemma 16, we will need the following precise version of Stirling's approximation.

Lemma 17 (Robbins 1955). *For each positive integer n , it holds that*

$$n! = \sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{r_n},$$

where $\frac{1}{12n+1} < r_n < \frac{1}{12n}$.

Lemma 17 yields a relatively simple formula for the probability mass function of the multinomial distribution.

Lemma 18. *For any $n \in \mathbb{N}$, $\mathbf{p} \in \Delta^{k-1}$, and $\mathbf{x} \in \Delta_n^{k-1}$ such that $x_i > 0$ for all $i \in [k]$, the probability $\text{Mult}(n, \mathbf{p}) \{\mathbf{x}\}$ is equal to*

$$\frac{1}{e^{r_{\mathbf{x}}}} \cdot \frac{\sqrt{n}}{(2\pi)^{\frac{k-1}{2}} \sqrt{\prod_{i \in [k]} x_i}} \cdot \frac{1}{e^{n \cdot D_{\text{KL}}(\frac{\mathbf{x}}{n} \| \mathbf{p})}},$$

where

$$0 \leq r_{\mathbf{x}} \leq \frac{k}{\min_{i \in [k]} x_i}.$$

Proof. Letting r_n be as in Lemma 17, we have

$$\text{Mult}(n, \mathbf{p}) \{\mathbf{x}\} = \binom{n}{\mathbf{x}} \prod_{i \in [k]} p_i^{x_i}$$

$$\begin{aligned} &= \frac{\sqrt{2\pi n} \left(\frac{n}{e} \right)^n e^{r_n}}{\prod_{i \in [k]} (\sqrt{2\pi x_i} \left(\frac{x_i}{e} \right)^{x_i} e^{r_{x_i}})} \prod_{i \in [k]} p_i^{x_i} \\ &= \frac{1}{e^{r_{\mathbf{x}}}} \cdot \frac{\sqrt{n}}{(2\pi)^{\frac{k-1}{2}} \sqrt{\prod_{i \in [k]} x_i}} \cdot \prod_{i \in [k]} \left(\frac{p_i}{x_i/n} \right)^{x_i} \\ &= \frac{1}{e^{r_{\mathbf{x}}}} \cdot \frac{\sqrt{n}}{(2\pi)^{\frac{k-1}{2}} \sqrt{\prod_{i \in [k]} x_i}} \cdot e^{-n \cdot D_{\text{KL}}(\frac{\mathbf{x}}{n} \| \mathbf{p})}, \end{aligned}$$

where $r_{\mathbf{x}} := \sum_{i \in [k]} r_{x_i} - r_n$. From the bounds of r_n in Lemma 17, it follows easily that $0 \leq r_{\mathbf{x}} \leq \frac{k}{\min_{i \in [k]} x_i}$. $\square$

Lemma 16 is then an immediate corollary of Lemma 18 (with $\mathbf{x} = \frac{n}{k} \cdot \mathbf{1}$ and $\tilde{r}_{n,k} = r_{\frac{n}{k} \cdot \mathbf{1}}$).

B.3 Proof of Lemma 4

The upper bound follows directly from Observation 2, Lemma 16, and non-negativity of the KL-divergence.

For the lower bound, observe that due to the symmetry over bundles when $A \in \mathcal{A}_{\text{bal}}$, we have $p_1^A = \dots = p_k^A = \frac{1}{k}$, and so $D_{\text{KL}}(\frac{1}{k} \| \mathbf{p}) = 0$. Again, Observation 2 and Lemma 16 then imply the bound.

B.4 Proof of Lemma 5

To prove Lemma 5, we will use the standard Hoeffding bound for sampling without replacement. The following lemma is stated as Corollary 1.1 in the work of Serfling (1974).¹⁴

Lemma 19 (Serfling 1974). *Let $X_1, \dots, X_t$ be random variables drawn from some list $x_1, \dots, x_T \in \{0, 1\}$ without replacement. Let $S = X_1 + \dots + X_t$ and $\mu = \mathbb{E}[S]$. For any $\theta > 0$, we have*

$$\Pr[S - \mu \geq \theta] \leq \exp(-2\theta^2/t);$$

$$\Pr[S - \mu \leq -\theta] \leq \exp(-2\theta^2/t).$$

Proof of Lemma 5. To compute the probability that $A^{(1)}, \dots, A^{(n^{2k})}$ are δ -IB, let us fix $A^{(v)}$ for some $v \in [n^{2k}]$. Then, consider $A^{(w)}$ for $w \in [n^{2k}] \setminus \{v\}$ that is picked at random, and any $i, j \in [k]$. Once $A_i^{(v)}$ is fixed, we let $x_g := \mathbf{1}[g \in A_i^{(v)}]$ for each $g \in M$, and view these values as in Lemma 19. Then, $A_j^{(w)} = \{g_1, \dots, g_{m/k}\}$ is chosen uniformly at random among all subsets of M of size m/k ; let $X_1 = x_{g_1}, \dots, X_{m/k} = x_{g_{m/k}}$. Observe that $S := X_1 + \dots + X_{m/k}$ is equal to $|A_i^{(v)} \cap A_j^{(w)}|$ and $\mathbb{E}[S] = m/k^2$. As such, we may apply Lemma 19 with $\theta = \frac{\delta m}{k^2}$ to conclude that

$$\begin{aligned} &\Pr \left[|A_i^{(v)} \cap A_j^{(w)}| \notin \left[\frac{m(1-\delta)}{k^2}, \frac{m(1+\delta)}{k^2} \right] \right] \\ &\leq 2 \exp(-2\delta^2 m/k^3) \\ &\leq \frac{2}{n^{8k}}, \end{aligned}$$

¹⁴In fact, Serfling's version is slightly stronger because it contains the term $1 - f_n^*$.

where the last inequality is due to $m \geq \frac{4k^4}{\delta^2} \cdot \ln n$.

Taking the union bound over all distinct $v, w \in [n^{2k}]$ and all $i, j \in [k]$, we have $\Pr[A^{(1)}, \dots, A^{(n^{2k})} \text{ are } \delta\text{-IB}] \geq 1 - \frac{2k^2 n^{4k}}{n^{8k}} = 1 - o(1)$ as desired. $\square$

B.5 Proof of Lemma 6

Let $\zeta = \frac{\beta}{4k}$ and $\gamma = \frac{\zeta}{k^2} = \frac{\beta}{4k^3}$. Applying Lemma 7 yields $\Pr[\mathcal{E}_{A'} \mid \mathcal{E}_A] \leq \Pr[X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}]$, where $X^1, \dots, X^n$ are as defined there. By Lemma 9, $X^1, \dots, X^n$ satisfy the condition of Lemma 10; in particular, since we assume that k is fixed and n grows, the condition $n \geq \frac{k}{1-2\zeta}$ is met. Hence,

$$\begin{aligned} \Pr[\mathcal{E}_{A'} \mid \mathcal{E}_A] &\leq \Pr\left[X^1 + \dots + X^n = \frac{n}{k} \cdot \mathbf{1}\right] \\ &\leq \exp(-2\zeta^2 n) + \frac{k^{k/2}}{(2\pi(1-2\zeta)n)^{\frac{k-1}{2}}} \\ &\leq o\left(\frac{1}{n^{\frac{k-1}{2}}}\right) + \frac{1}{(1-2\zeta)^k} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}} \\ &\leq o\left(\frac{1}{n^{\frac{k-1}{2}}}\right) + \frac{1}{1-\beta/2} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}} \\ &\leq \frac{1}{1-\beta/2-o(1)} \cdot \frac{k^{k/2}}{(2\pi n)^{\frac{k-1}{2}}}, \end{aligned}$$

where we use Bernoulli's inequality for the penultimate inequality.

B.6 Proof of Lemma 9

To prove Lemma 9, we need the following bound on the total variation distance between two centered Gaussians, stated as Theorem 1.1 in the work of Devroye, Mehrabian, and Reddad (2023).

Lemma 20 (Devroye, Mehrabian, and Reddad 2023). *For positive definite matrices $\Sigma, \tilde{\Sigma} \in \mathbb{R}^{d \times d}$, let $\lambda_1, \dots, \lambda_d$ be the eigenvalues of $\tilde{\Sigma}^{-1}\Sigma - I_d$. Then,*

$$D_{\text{TV}}\left(\mathcal{N}(\mathbf{0}, \Sigma), \mathcal{N}(\mathbf{0}, \tilde{\Sigma})\right) \leq \frac{3}{2} \sqrt{\sum_{i=1}^d \lambda_i^2}.$$

We are now ready to show that the probabilities $p_{(i,i)}$ are all close to $1/k^2$.

Proof of Lemma 9. Let μ and σ^2 denote the mean and variance of $\mathcal{D}$, respectively. For each $g \in M$, let $W^g \in \mathbb{R}^{k^2}$ be a random variable such that

$$W_{(j,j')}^g = \begin{cases} (u_a(g) - \mu) \cdot \frac{k}{\sigma\sqrt{m}} & \text{if } g \in A_j \cap A_{j'}; \\ 0 & \text{otherwise} \end{cases}$$

for all $j, j' \in [k]$; note that this definition is the same regardless of a . Observe also that the random variables W^g are independent for different g (since the utilities are independent) and have mean zero. Next, let $H = \sum_{g \in M} W^g$. It

can be seen that¹⁵ $\text{Cov}[H] = \text{diag}((\sigma_{(j,j')}^2)_{j,j' \in [k]})$, where $\sigma_{(j,j')} = \sqrt{\frac{k^2}{m} \cdot |A_j \cap A_{j'}|}$; we write Σ as a shorthand for $\text{diag}((\sigma_{(j,j')}^2)_{j,j' \in [k]})$. Finally, let $\Lambda_{i,i'}$ denote all vectors $\mathbf{v} \in \mathbb{R}^{k^2}$ that satisfy

$$\sum_{\ell \in [k]} v_{(i,\ell)} \geq \sum_{\ell \in [k]} v_{(j,\ell)} \text{ and } \sum_{\ell \in [k]} v_{(\ell,i')} \geq \sum_{\ell \in [k]} v_{(\ell,j')}$$

for all $j, j' \in [k]$. Since $\Lambda_{i,i'}$ is convex, we can apply Lemma 8 to get

$$\begin{aligned} &|\Pr[H \in \Lambda_{i,i'}] - \mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\}| \\ &\leq 60k^{1/2} \sum_{g \in M} \mathbb{E}[\|\Sigma^{-1/2}W^g\|_2^3]. \end{aligned}$$

Now, notice that for any $j, j' \in [k]$,

$$\begin{aligned} &(\Sigma^{-1/2}W^g)_{(j,j')} \\ &= \begin{cases} (u_a(g) - \mu) \cdot \frac{1}{\sigma\sqrt{|A_j \cap A_{j'}|}} & \text{if } g \in A_j \cap A_{j'}; \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Thus, since $u_a(g) \in [0, 1]$, we have $\|\Sigma^{-1/2}W^g\|_2 \leq \frac{1}{\sigma\sqrt{|A_j \cap A_{j'}|}}$ for $g \in A_j \cap A_{j'}$. Plugging this into the inequality above yields

$$\begin{aligned} &|\Pr[H \in \Lambda_{i,i'}] - \mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\}| \\ &\leq 60k^{1/2} \sum_{j,j' \in [k]} \sum_{g \in A_j \cap A_{j'}} \frac{1}{\sigma^3 |A_j \cap A_{j'}|^{1.5}} \\ &= 60k^{1/2} \sum_{j,j' \in [k]} \frac{1}{\sigma^3 \sqrt{|A_j \cap A_{j'}|}} \\ &\leq 60k^{1/2} \sum_{j,j' \in [k]} \frac{1}{\sigma^3 \sqrt{\frac{m(1-\delta)}{k^2}}} \\ &= \sqrt{\frac{3600k^7}{\sigma^6} \cdot \frac{1}{m(1-\delta)}} \\ &\leq 0.5\gamma, \end{aligned} \tag{1}$$

where the second inequality is due to our assumption that A, A' are δ -IB and the last inequality follows from our assumption on m and k .

Next, we bound the TV distance between $\mathcal{N}(\mathbf{0}, \Sigma)$ and $\mathcal{N}(\mathbf{0}, I_{k^2})$ via Lemma 20. To this end, observe that $\Sigma - I_{k^2} = \text{diag}\left(\left(\frac{k^2}{m} \cdot |A_j \cap A_{j'}| - 1\right)_{j,j' \in [k]}\right)$. Thus, the eigenvalues of $\Sigma - I_d$ are $\frac{k^2}{m} \cdot |A_j \cap A_{j'}| - 1$ for all $j, j' \in [k]$. Since A, A' are δ -IB, we have

$$\left(\frac{k^2}{m} \cdot |A_j \cap A_{j'}| - 1\right)^2 \leq \delta^2.$$

¹⁵Recall that, for $\mathbf{u} \in \mathbb{R}^d$, $\text{diag}(\mathbf{u}) \in \mathbb{R}^{d \times d}$ denotes the diagonal matrix whose diagonal entries are $\mathbf{u}$ and remaining entries are zero.

Plugging this into Lemma 20, we conclude that

$$D_{\text{TV}}(\mathcal{N}(\mathbf{0}, \Sigma), \mathcal{N}(\mathbf{0}, I_{k^2})) \leq \frac{3}{2} \cdot k\delta \leq 0.5\gamma. \quad (2)$$

Therefore, by the triangle inequality, we have

$$\begin{aligned} & |\Pr[H \in \Lambda_{i,i'}] - \mathcal{N}(\mathbf{0}, I_{k^2})\{\Lambda_{i,i'}\}| \\ & \leq |\Pr[H \in \Lambda_{i,i'}] - \mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\}| \\ & \quad + |\mathcal{N}(\mathbf{0}, \Sigma)\{\Lambda_{i,i'}\} - \mathcal{N}(\mathbf{0}, I_{k^2})\{\Lambda_{i,i'}\}| \stackrel{(1),(2)}{\leq} \gamma. \end{aligned}$$

Observe that by definition of $\Lambda_{i,i'}$ and non-atomicity of $\mathcal{D}$, it holds that $p_{(i,i')} = \Pr[H \in \Lambda_{i,i'}]$. Moreover, due to symmetry, we have

$$\mathcal{N}(\mathbf{0}, I_{k^2})\{\Lambda_{i,i'}\} = \mathcal{N}(\mathbf{0}, I_{k^2})\{\Lambda_{j,j'}\}$$

for all $j, j' \in [k]$. Note that $\bigcup_{j,j' \in [k]} \Lambda_{j,j'} = \mathbb{R}^{k^2}$ and the intersection of any two distinct sets $\Lambda_{j,j'}$ has measure zero (with respect to $\mathcal{N}(\mathbf{0}, I_{k^2})$). As such, we have that $\mathcal{N}(\mathbf{0}, I_{k^2})\{\Lambda_{i,i'}\} = \frac{1}{k^2}$. Combining this with the inequality above concludes our proof. $\square$

B.7 Proof of Lemma 10

We introduce an additional notation. For any random variable X , let $\mathfrak{d}(X) = \max_{x \in \text{supp}(X)} \Pr[X = x]$. Sometimes we also write the distribution of X instead of X itself. We use the following properties of $\mathfrak{d}$, whose proofs we defer to Section B.8.

Lemma 21. *For independent $X, Y \in \mathbb{Z}^d$, it holds that $\mathfrak{d}(X + Y) \leq \mathfrak{d}(X)$.*

Lemma 22. *For any positive integer $r \geq k$, the following inequality holds: $\mathfrak{d}(\text{Mult}(r, \frac{1}{k} \cdot \mathbf{1})) \leq \frac{k^{k/2}}{(2\pi r)^{\frac{k-1}{2}}}$.*

We also need Hoeffding's inequality, which is stated in terms of the binomial distribution for convenience.

Lemma 23 (Hoeffding's inequality). *For $\psi \in (0, 1)$, $t > 0$, and $n \in \mathbb{N}$, $\Pr_{S \sim \text{Bin}(n, \psi)}[S < n\psi - t] \leq \exp(-2t^2/n)$.*

We wish to upper bound $\Pr[\tilde{X}^1 + \dots + \tilde{X}^n = \frac{n}{k} \cdot \mathbf{1}]$.

By definition, this is at most $\mathfrak{d}(\tilde{X}^1 + \dots + \tilde{X}^n)$. To bound

this quantity, we reinterpret $\tilde{X}^a$ so that it is a mixture distribution between the uniform distribution on $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ and a “leftover” distribution. In other words, for each $\tilde{X}^a$, we can toss a (biased) coin and, based on the outcome, sample $\tilde{X}^a$ either from the uniform distribution or from the leftover distribution. Using Lemma 23, we show that $\tilde{X}^a$ is drawn from the uniform distribution for most a . For these $\tilde{X}^a$, their sum exactly follows the multinomial distribution, so we can apply Lemma 22. Finally, Lemma 21 allows us to handle the leftover.

Proof of Lemma 10. It suffices to bound $\mathfrak{d}(\tilde{X}^1 + \dots + \tilde{X}^n)$.

Instead of sampling each $\tilde{X}^a$ directly, we can alternatively sample $\tilde{X}^a$ as follows.

- Sample $Z^a \sim \text{Ber}(1 - \zeta)$.
- Sample U^a uniformly at random from $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$.

- Sample V^a from $\{\mathbf{e}_1, \dots, \mathbf{e}_k\}$ where $\Pr[V^a = \mathbf{e}_i] = \frac{1}{\zeta} \left(\Pr[\tilde{X}^a = \mathbf{e}_i] - (1 - \zeta)\frac{1}{k} \right)$.
- If $Z^a = 1$, let $\tilde{X}^a = U^a$; if $Z^a = 0$, let $\tilde{X}^a = V^a$.

Observe that V^a is a valid probability distribution. Also, one can check that the process defined above results in the correct probability distribution for $\tilde{X}^a$.

Therefore, $\mathfrak{d}(\tilde{X}^1 + \dots + \tilde{X}^n)$ can be written as

$$\begin{aligned} & \max_{\mathbf{x} \in \Delta_n^{k-1}} \Pr[\tilde{X}^1 + \dots + \tilde{X}^n = \mathbf{x}] \\ & = \max_{\mathbf{x} \in \Delta_n^{k-1}} \sum_{S \subseteq N} \Pr \left[\sum_{a \in S} U^a + \sum_{a' \notin S} V^{a'} = \mathbf{x} \right] \\ & \quad \cdot \Pr[S = \{a \in N \mid Z^a = 1\}] \\ & \leq \sum_{S \subseteq N} \mathfrak{d} \left(\sum_{a \in S} U^a + \sum_{a' \notin S} V^{a'} \right) \\ & \quad \cdot \Pr[S = \{a \in N \mid Z^a = 1\}] \\ & \stackrel{(\text{Lemma 21})}{\leq} \sum_{S \subseteq N} \mathfrak{d} \left(\sum_{a \in S} U^a \right) \\ & \quad \cdot \Pr[S = \{a \in N \mid Z^a = 1\}]. \end{aligned}$$

Observe that $\sum_{a \in S} U^a$ is distributed as $\text{Mult}(|S|, \frac{1}{k} \cdot \mathbf{1})$ and that $|S| = |\{a \in N \mid Z^a = 1\}|$ is distributed as $\text{Bin}(n, 1 - \zeta)$. Thus, we can write the above inequality as

$$\begin{aligned} & \mathfrak{d}(\tilde{X}^1 + \dots + \tilde{X}^n) \\ & \leq \mathbb{E}_{r \sim \text{Bin}(n, 1 - \zeta)} \left[\mathfrak{d} \left(\text{Mult} \left(r, \frac{1}{k} \cdot \mathbf{1} \right) \right) \right] \\ & \leq \Pr_{r \sim \text{Bin}(n, 1 - \zeta)}[r \leq (1 - 2\zeta)n] \\ & \quad + \mathbb{E}_{r \sim \text{Bin}(n, 1 - \zeta)} \left[\mathfrak{d} \left(\text{Mult} \left(r, \frac{1}{k} \cdot \mathbf{1} \right) \right) \mid r > (1 - 2\zeta)n \right], \end{aligned}$$

where the last inequality follows from the law of total expectation. Applying Lemmas 23 and 22 respectively to bound the two remaining terms completes the proof. $\square$

B.8 Properties of $\mathfrak{d}$: Proofs of Lemmas 21 and 22

Proof of Lemma 21. For any $z \in \text{supp}(X + Y)$, we have

$$\begin{aligned} \Pr[X + Y = z] & = \sum_{y \in \text{supp}(Y)} \Pr[X = z - y] \Pr[Y = y] \\ & \leq \sum_{y \in \text{supp}(Y)} \mathfrak{d}(X) \Pr[Y = y] = \mathfrak{d}(X). \end{aligned}$$

This implies that $\mathfrak{d}(X + Y) \leq \mathfrak{d}(X)$. $\square$

For Lemma 22, we will use the following properties of the gamma function. These properties are well-known and their proofs can be found in standard textbooks on the topic (e.g., Artin 1964). Recall that a function is said to be logarithmically convex if its logarithm is convex.

Lemma 24. $\Gamma : \mathbb{R} \rightarrow \mathbb{R}$ is a function that satisfies

- (i) $\Gamma(1+r) = r!$ for all $r \in \mathbb{Z}_{\geq 0}$;
- (ii) Γ is logarithmically convex on $(0, \infty)$;
- (iii) For $x \geq 1$, $\Gamma(1+x) = \sqrt{2\pi x}(x/e)^x e^{\mu(x)}$, where $\mu : [1, \infty) \rightarrow \mathbb{R}_{\geq 0}$ is decreasing.

Proof of Lemma 22. Consider any $\mathbf{x} \in \Delta_r^{k-1}$, we have

$$\begin{aligned}
\text{Mult}\left(r, \frac{1}{k} \cdot \mathbf{1}\right)\{\mathbf{x}\} &= \frac{r!}{\prod_{i \in [k]} x_i!} \cdot \frac{1}{k^r} \\
&= \frac{\Gamma(r+1)}{\prod_{i \in [k]} \Gamma(x_i+1)} \cdot \frac{1}{k^r} \\
&\leq \frac{\Gamma(r+1)}{\Gamma(r/k+1)^k} \cdot \frac{1}{k^r} \\
&\leq \frac{\sqrt{2\pi r}(r/e)^r}{\left(\sqrt{2\pi r/k} \left(\frac{r}{ke}\right)^{r/k}\right)^k} \cdot \frac{1}{k^r} \\
&= \frac{k^{k/2}}{(2\pi r)^{\frac{k-1}{2}}},
\end{aligned}$$

where we use the properties of the gamma function from Lemma 24 in the respective order. $\square$

We remark that bounds in the same vein as Lemma 10 have been shown, e.g., by Ushakov (1986) and Postnikov and Yudin (1988). However, these are not sufficient for our purposes, since either they require the distributions to be identical or the provided probability bound is not tight for $k > 2$ (up to a constant factor).

B.9 Proof of Lemma 11

The proof of Lemma 11 requires a few intermediate lemmas. To state these lemmas, we need an additional notation. For any non-atomic real-valued distribution $\mathcal{P}$ and any $p \in (0, 1)$, let $Q_{\mathcal{P}}(p)$ denote its p -th quantile, i.e., the (unique) point x such that $\Pr_{X \sim \mathcal{P}}[X < x] = p$.

We start with the following inequality, which is an immediate consequence of Chebyshev's inequality.

Lemma 25. *For any non-atomic real-valued distribution $\mathcal{P}$ with mean μ and standard deviation $\sigma_{\mathcal{P}}$ and any $p \in (0, 1)$,*

$$\mu - \frac{\sigma_{\mathcal{P}}}{\sqrt{p}} \leq Q_{\mathcal{P}}(p) \leq \mu + \frac{\sigma_{\mathcal{P}}}{\sqrt{1-p}}.$$

Next is a bound on the variance of a distribution after conditioning it to be larger than a certain value.

Lemma 26. *For any non-atomic real-valued distribution $\mathcal{P}$ with standard deviation $\sigma_{\mathcal{P}}$ and any $p \in (0, 1)$, let $\mathcal{P}_{\geq x}$ denote the conditional distribution of $\mathcal{P}$ on $[x, \infty)$. Then, we have $\text{Var}(\mathcal{P}_{\geq Q_{\mathcal{P}}(p)}) \leq \sigma_{\mathcal{P}}^2/(1-p)$.*

Proof. For any distribution $\mathcal{Y}$ with mean $\mu_{\mathcal{Y}}$ and any real number c , one can check that $\mathbb{E}_{Y \sim \mathcal{Y}}[(Y-c)^2] = \text{Var}(\mathcal{Y}) + (c - \mu_{\mathcal{Y}})^2 \geq \text{Var}(\mathcal{Y})$.

Let $\mu_{\mathcal{P}}$ denote the mean of $\mathcal{P}$. We have

$$\begin{aligned}
\text{Var}(\mathcal{P}_{\geq Q_{\mathcal{P}}(p)}) \\
\leq \mathbb{E}_{X \sim \mathcal{P}_{\geq Q_{\mathcal{P}}(p)}}[(X - \mu_{\mathcal{P}})^2]
\end{aligned}$$

$$\begin{aligned}
&= \frac{\mathbb{E}_{X \sim \mathcal{P}}[(X - \mu_{\mathcal{P}})^2 \cdot \mathbf{1}[X \geq Q_{\mathcal{P}}(p)]]}{1-p} \\
&\leq \frac{\mathbb{E}_{X \sim \mathcal{P}}[(X - \mu_{\mathcal{P}})^2]}{1-p} \\
&= \frac{\sigma_{\mathcal{P}}^2}{1-p},
\end{aligned}$$

as desired. $\square$

We next bound the probability that two independent and identically distributed (i.i.d.) random variables differ by at most a certain amount.

Lemma 27. *For any non-atomic real-valued distribution $\mathcal{P}'$ with standard deviation σ' and any $\xi > 0$, we have*

$$\Pr_{Y_1, Y_2 \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}'}[|Y_1 - Y_2| < \xi] \geq \min\left\{\frac{1}{8}, \frac{\xi}{32\sigma'}\right\}.$$

Proof. We may assume that $\xi \leq 4\sigma'$, as the right-hand side does not increase further if $\xi > 4\sigma'$. Lemma 25 implies that $|Q_{\mathcal{P}'}(0.25) - Q_{\mathcal{P}'}(0.75)| \leq 4\sigma'$. Let $T = \lceil 4\sigma'/\xi \rceil \geq 1$ and $I_1, \dots, I_T$ be any partition of $[Q_{\mathcal{P}'}(0.25), Q_{\mathcal{P}'}(0.75)]$ into disjoint intervals of length at most ξ each. We have

$$\begin{aligned}
&\Pr_{Y_1, Y_2 \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}'}[|Y_1 - Y_2| < \xi] \\
&\geq \sum_{t \in [T]} \Pr_{Y_1, Y_2 \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}'}[Y_1, Y_2 \in I_t] \\
&= \sum_{t \in [T]} \Pr_{Y \sim \mathcal{P}'}[Y \in I_t]^2 \\
&\stackrel{(*)}{\geq} \frac{1}{T} \left(\sum_{t \in [T]} \Pr_{Y \sim \mathcal{P}'}[Y \in I_t] \right)^2 \\
&= \frac{1}{T} \cdot \Pr_{Y \sim \mathcal{P}'}[Y \in [Q_{\mathcal{P}'}(0.25), Q_{\mathcal{P}'}(0.75)]]^2 \\
&= \frac{1}{4T} \\
&\geq \frac{\xi}{32\sigma'},
\end{aligned}$$

where $(*)$ follows from the Cauchy-Schwarz inequality and the last inequality from our choice of T . $\square$

Using the two bounds above, we now derive an inequality concerning the difference between the maximum and a non-maximum among k i.i.d. random variables.

Lemma 28. *Let $\mathcal{P}$ be any non-atomic real-valued distribution with standard deviation $\sigma_{\mathcal{P}}$. Let $Z_1, Z_2, \dots, Z_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}$. Then, for any $\xi > 0$,*

$$\begin{aligned}
&\Pr[Z_1 > Z_2, \dots, Z_k \text{ and } Z_1 - Z_k \leq \xi] \\
&\geq \frac{1}{16k^2} \cdot \min\left\{\frac{1}{8}, \frac{\xi}{64\sigma_{\mathcal{P}}\sqrt{k}}\right\}.
\end{aligned}$$

Proof. By symmetry, we have

$$\Pr[Z_1 > Z_2, \dots, Z_k \text{ and } Z_1 - Z_k \leq \xi]$$

$$\begin{aligned}
&= \frac{1}{2} \left(\Pr[Z_1 > Z_2, \dots, Z_k \text{ and } Z_1 - Z_k \leq \xi] \right. \\
&\quad \left. + \Pr[Z_k > Z_1, \dots, Z_{k-1} \text{ and } Z_k - Z_1 \leq \xi] \right) \\
&\geq \frac{1}{2} \cdot \Pr[Z_1, Z_k > Z_2, \dots, Z_{k-1} \text{ and } |Z_1 - Z_k| \leq \xi].
\end{aligned}$$

To bound this term further, let $p = 1 - \frac{1}{2k}$ and $q = Q_{\mathcal{P}}(p)$. We have

$$\begin{aligned}
&\Pr[Z_1, Z_k > Z_2, \dots, Z_{k-1} \text{ and } |Z_1 - Z_k| \leq \xi] \\
&\geq \Pr[Z_1, Z_k \geq q > Z_2, \dots, Z_{k-1} \text{ and } |Z_1 - Z_k| \leq \xi] \\
&= (1-p)^2 p^{k-2} \cdot \Pr[|Z_1 - Z_k| \leq \xi \mid Z_1, Z_k \geq q] \\
&= (1-p)^2 p^{k-2} \cdot \Pr_{Z'_1, Z'_k \stackrel{\text{i.i.d.}}{\sim} \mathcal{P}_{\geq q}}[|Z'_1 - Z'_k| \leq \xi] \\
&\geq (1-p)^2 p^{k-2} \cdot \min \left\{ \frac{1}{8}, \frac{\xi \sqrt{1-p}}{32\sigma_{\mathcal{P}}} \right\},
\end{aligned}$$

where the last inequality follows from Lemmas 26 and 27.

Finally, combining the two inequalities above and using the definition of p , we have

$$\begin{aligned}
&\Pr[Z_1 > Z_2, \dots, Z_k \text{ and } Z_1 - Z_k \leq \xi] \\
&\geq \frac{1}{2} \cdot \frac{1}{4k^2} \cdot \left(1 - \frac{1}{2k}\right)^{k-2} \cdot \min \left\{ \frac{1}{8}, \frac{\xi}{64\sigma_{\mathcal{P}}\sqrt{k}} \right\} \\
&\geq \frac{1}{16k^2} \cdot \min \left\{ \frac{1}{8}, \frac{\xi}{64\sigma_{\mathcal{P}}\sqrt{k}} \right\},
\end{aligned}$$

where the last inequality is due to Bernoulli's inequality. $\square$

We are now ready to prove Lemma 11. At a high level, we first recall that in the case where A is balanced, each p_i^A is equal to $1/k$. We then analyze the “probability deficit” that p_i^A incurs when a good is added to some other bundle $A_{i'}$. Lemma 28 ensures that, before the addition, the difference between $u_a(A_i)$ and $u_a(A_{i'})$ is small. This allows us to lower-bound the probability deficit, which is incurred when the additional value exceeds the difference.

Proof of Lemma 11. Assume without loss of generality that $|A_1| \leq |A_2| \leq \dots \leq |A_k|$. We will show that the desired inequality holds for $i = 1$. The case $|A_1| = 0$ is trivial, so we henceforth assume $|A_1| \geq 1$. Let $A'_2, \dots, A'_k$ be arbitrary subsets of $A_2, \dots, A_k$ of size $|A_1|$, respectively, and let $Z_1 = u_a(A_1), Z_2 = u_a(A'_2), \dots, Z_k = u_a(A'_k)$. Since $Z_1, \dots, Z_k$ are i.i.d. and non-atomic, we have

$$\Pr[Z_1 \geq Z_2, \dots, Z_k] = \frac{1}{k}.$$

Thus, we can bound the desired probability as

$$\begin{aligned}
&\Pr[u_a(A_1) \geq u_a(A_2), \dots, u_a(A_k)] \\
&\leq \Pr[u_a(A_1) \geq u_a(A'_2), \dots, u_a(A'_{k-1}), u_a(A_k)] \\
&= \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } u_a(A_k) \leq Z_1] \\
&= \Pr[Z_1 \geq Z_2, \dots, Z_k] \\
&\quad - \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } u_a(A_k) > Z_1] \\
&= \frac{1}{k} - \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } u_a(A_k) > Z_1].
\end{aligned}$$

Now, observe that $u_a(A_k) = Z_k + u_a(A_k \setminus A'_k)$. Since m is not divisible by k , we have $A_k \setminus A'_k \neq \emptyset$. By definition of quantiles, it holds that $\Pr[u_a(A_k \setminus A'_k) > Q_{\mathcal{D}}(1/2)] \geq 1/2$. Moreover, $u_a(A_k \setminus A'_k)$ is independent of $Z_1, \dots, Z_k$. Hence, we have

$$\begin{aligned}
&\Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } u_a(A_k) > Z_1] \\
&= \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } Z_1 - Z_k < u_a(A_k \setminus A'_k)] \\
&\geq \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } \\
&\quad Z_1 - Z_k < Q_{\mathcal{D}}(1/2) < u_a(A_k \setminus A'_k)] \\
&\geq \frac{1}{2} \cdot \Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } Z_1 - Z_k < Q_{\mathcal{D}}(1/2)].
\end{aligned}$$

For any random variable Y supported on $[0, 1]$, it holds that

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \leq \mathbb{E}[Y] - \mathbb{E}[Y]^2 \leq 1/4.$$

Since each of $Z_1, \dots, Z_k$ is a sum of $|A_1|$ independent random variables supported on $[0, 1]$, we have

$$\text{Var}(Z_1), \dots, \text{Var}(Z_k) \leq \frac{|A_1|}{4} \leq \frac{m}{4k}.$$

Moreover, $Q_{\mathcal{D}}(1/2) > 0$ since $\mathcal{D}$ is non-atomic and supported on $[0, 1]$. Thus, we may apply Lemma 28 to get

$$\begin{aligned}
&\Pr[Z_1 \geq Z_2, \dots, Z_k \text{ and } Z_1 - Z_k < Q_{\mathcal{D}}(1/2)] \\
&\geq \frac{Q_{\mathcal{D}}(1/2)}{512k^2\sqrt{m}}.
\end{aligned}$$

Finally, putting everything together yields the desired bound for $\alpha = \frac{Q_{\mathcal{D}}(1/2)}{1024}$. $\square$

B.10 Proof of Lemma 12

From Observation 2, Lemma 16, and the fact that $n \geq k$, we have $\Pr[\mathcal{E}_A] \leq \frac{\sqrt{k}}{e^{n \cdot \text{D}_{\text{KL}}(\frac{1}{k} \| \mathbf{p}^A)}} \cdot \text{Also, Lemmas 15 and 11}$

respectively imply that $\text{D}_{\text{KL}}(\frac{1}{k} \| \mathbf{p}^A) \geq 2 \cdot \text{D}_{\text{TV}}(\frac{1}{k}, \mathbf{p}^A)^2 \geq \frac{\alpha^2}{k^4 m}$. Combining these yields the claimed bound.