Robust Rent Division

Dominik Peters[1,2], Ariel D. Procaccia[1], and David Zhu[1]

[1]Harvard University
[2]CNRS, LAMSADE, Université Paris Dauphine–PSL

October 2022

In fair rent division, the problem is to assign rooms to roommates and fairly split the rent based on roommates' reported valuations for the rooms. Envy-free rent division is the most popular application on the fair division website Spliddit. The standard model assumes that agents can correctly report their valuations for each room. In practice, agents may be unsure about their valuations, for example because they have had only limited time to inspect the rooms. Our goal is to find a robust rent division that remains fair even if agent valuations are slightly different from the reported ones. We introduce the lexislack solution, which selects a rent division that remains envy-free for valuations within as large a radius as possible of the reported valuations. We also consider robustness notions for valuations that come from a probability distribution, and use results from learning theory to show how we can find rent divisions that (almost) maximize the probability of being envy-free, or that minimize the expected envy. We show that an almost optimal allocation can be identified based on polynomially many samples from the valuation distribution. Finding the best allocation given these samples is NP-hard, but in practice such an allocation can be found using integer linear programming.

1. Introduction

The literature on fair division of resources has produced allocation mechanisms for many domains, such as course allocation, indivisible goods, chores, house assignment, and the selection of citizens' assemblies [Budish, 2011, Caragiannis et al., 2019, Moulin, 2019, Flanigan et al., 2021]. But arguably the most widely *used* example is *rent division*: this is the most popular application on the fair division website spliddit.org [Goldman and Procaccia, 2014], where it has been used more than 30,000 times since its launch in 2014.

Rent division deals with the common situation where a group of n future roommates are planning to move into a house or apartment which has n rooms, one for each roommate. They will split the rent payments among themselves. The roommates may differ in how much they are willing to pay for different rooms. Given the room valuations of each roommate, our task is to assign the rooms, and to decide how to split the rent. We wish to do this fairly, and so we will choose an allocation that is *envy-free*: no roommate would strictly prefer to get another room, given the prices we have assigned to those rooms. Such an allocation is guaranteed to exist [Svensson, 1983].

Let us consider an example with $n = 3$ roommates, and let the total rent be \$1000. Table 1 shows the valuation that each agent assigns to each room. Given this information, the algorithm in use on Spliddit will assign Room 1 to Alice, Room 2 to Bob, and Room 3 to Charlie, charging them \$100, \$500, and \$400 respectively. This allocation is envy-free under the assumption (which we will make throughout) that agents have *quasilinear utilities*: their utility under an allocation is the value of their room minus its price. For example, Alice has utility $300 - 100 = 200$. She does not envy the others: Bob's room would give her only utility $400 - 500 < 200$, and Charlie's room $300 - 400 < 200$.

On a typical instance, there are infinitely many allocations that are envy-free. Spliddit's algorithm chooses the one that maximizes the utility of the worst-off agent, subject to envy-freeness [Gal et al., 2017]. This is known as the *maximin* rule. In optimizing this objective, Spliddit might choose an outcome that is only barely envy-free. In the example, Bob has utility $700 - 500 = 200$, but he would gain the same utility from having Alice's room: $300 - 100$. If, upon moving in, Bob discovers a defect in his room and

	Room 1	Room 2	Room 3
Alice	**300**	400	300
Bob	300	**700**	0
Charlie	300	100	**600**
Spliddit	100	500	400
Lexislack	200	450	350

Table 1: Example of valuations. In any envy-free allocation, Alice gets room 1, Bob gets room 2, and Charlie gets room 3. The lower rows display the price vectors selected by Spliddit's rule (maximin) and by our lexislack rule.

now only values it at 600, say, then he would envy Alice. Thus, the envy-freeness of Spliddit's allocation is not robust.

We study the rent division problem with the goal of finding allocations that are robustly envy-free, in the sense that they remain envy-free even if valuations change slightly. For this, we introduce the *lexislack rule*, which selects an envy-free allocation where the minimum "slack" (the amount by which agent i prefers her allocation to agent j's) is maximized lexicographically. This produces an allocation that remains envy-free for all valuation profiles that are within a maximally large ℓ_1-radius of the reported profile. In the example of Table 1, the lexislack rule assigns the rooms in the same way as does Spliddit, but charges the roommates \$200, \$450, and \$350. One can verify that with these prices, each agent prefers their allocation to any other agent's by at least 150. This means that even after Bob's adjustment to 600, he does not envy Alice. We show that the lexislack rule always selects an essentially unique outcome, which can be found in polynomial time by linear programming.

This notion of robustness may not always be appropriate. Consider two perturbations with equal ℓ_1-distance to the reported valuations: one changes agent i's valuations for all rooms by a small amount, the other changes i's valuation for one room by a large amount. Lexislack places equal importance on them. But the former perturbation seems more likely: even if a player is uncertain about the value of a room, that value is more likely to be close to their best estimate than further away. Thus, arguably, different valuation profiles should be weighted differently: we do not want to sacrifice envy-freeness for a likely perturbation in order to obtain it for an unlikely perturbation.

To capture this idea, we propose to add noise — such as Gaussian noise — around the reported valuations. This way, we impute a probability distribution $\mathcal{D}$ over valuations. In this setting, our interpretation of robustness is to look for allocations that are envy-free with *maximum probability*. However, it is not clear how one could efficiently find the most robust allocation given the noisy valuations. As part of our methodological contribution, we propose an approach based on synthetic sampling. Specifically, we sample a number of valuation profiles from $\mathcal{D}$, and then find an allocation that is optimal on this sample using integer linear programming (ILP). By calculating the VC dimension of the space of rent divisions, we give polynomial sample complexity bounds that show how many samples are sufficient so that this approach identifies an almost optimal allocation with high probability. Note that the samples are synthetic, but low sample complexity is crucial nevertheless: a small number of samples leads to a sufficiently small ILP that, as we show, can be optimally solved in practice (even though we prove that the problem is NP-complete).

We also show that one can use the sampling approach to find an allocation that minimizes the expected amount by which one agent envies another. In contrast to maximizing the probability of envy-freeness, the minimum envy objective places more emphasis on avoiding bad violations of envy-freeness.

An advantage of our sampling-based approach is that it is very general and does not place any restrictions on the distribution $\mathcal{D}$. Our algorithms could also be used for rent division problems with uncertainty, where agents might explicitly report distributions over their valuations. For example, a simple Spliddit-like user interface could allow agents to report their room valuations as a *range* rather than a number.

We end with some experiments on data taken from Spliddit. They suggest that our three new rules significantly outperform the Spliddit maximin rule on robustness metrics. Interestingly, the lexislack solution does comparably well to the rules based on sampling. Given its conceptual simplicity and easy computation, this suggests lexislack as a good rule when robustness is desired.

Related Work

The rent division model is well-studied in the economics literature [Svensson, 1983, Alkan et al., 1991, Aragones, 1995, Su, 1999, Velez, 2018], often without assuming quasilinear utilities. That literature includes results on the structure of the envy-free set and about strategic aspects. Computer scientists have studied the computation of allocation rules [Gal et al., 2017, Procaccia et al., 2018]. Bei et al. [2021] study a generalization of the rent division problem.

Robustness has been studied in several areas of computational social choice, such as in voting [Shiryaev et al., 2013], in committee elections [Bredereck et al., 2021, Gawron and Faliszewski, 2019, Misra and Sonar, 2019], and in stable matching [Chen et al., 2019, Mai and Vazirani, 2018]. We are not aware of such work for fair division, though Menon and Larson [2020] study a related problem of "stability" which requires that the allocation should not change much if valuations change slightly. For rent division, a blog post by Critch [2015] argues in favor of aiming for robustness in the rent division problem. Critch [2015] implemented an algorithm for robust rent division that appears in experiments to maximize the slack, but it differs from the lexislack rule, and no theoretical analysis of this algorithm is available.

Our sampling-based approach is conceptually related to work on data-driven algorithm design [Balcan, 2020], which typically seeks to optimize the hyperparameters of an algorithm with respect to an underlying distribution over instances, based on samples. One thing that distinguishes our distributional setting is that we are using the samples to optimize a single solution to our problem. Computational hardness results for problems similar to our sample-based optimization problems have been obtained in the settings of stable matching and of Pareto-optimal assignment [Aziz et al., 2019, 2020].

2. Preliminaries: Rent Division

Let $n \in \mathbb{N}$ and write $[n] = \{1, \dots, n\}$. Let $N = [n]$ be a set of n agents, and let $R = [n]$ be a set of n rooms. Without loss of generality, we let the total rent be 1. A *(valuation) profile* $v = (v_{ir})_{i \in N, r \in R}$ is a collection of values $v_{ir} \in \mathbb{Q}_+$, one for each agent $i \in N$ and each room $r \in R$.

A *room assignment* is a bijection $\sigma : N \to R$, so that agent i is assigned room $\sigma(i)$. Given a valuation profile v, we say that σ is *optimal* if it maximizes utilitarian social welfare $\sum_{i \in N} v_{i\sigma(i)}$. An *allocation* (σ, p) is a room assignment σ together with a payment vector $p = (p_1, \dots, p_n) \in \mathbb{R}^n$ with $\sum_{r \in R} p_r = 1$, where p_r is the rent of room r. (The value p_r is usually non-negative.)

We assume that agents have *quasilinear utilities*. This means that if v is a valuation profile and (σ, p) is an allocation, then agent i's utility in this allocation is $v_{i\sigma(i)} - p_{\sigma(i)}$, i.e., the valuation of i for her room $\sigma(i)$ minus the room's rent. An allocation (σ, p) is *envy-free* if $v_{i\sigma(i)} - p_{\sigma(i)} \geqslant v_{ir} - p_r$ for all $i \in N$ and $r \in R$, so that each agent i weakly prefers her allocation to receiving any other room.

A *solution* is a function that given a valuation profile, selects a set of allocations (usually a singleton, but ties may occur). A solution is *essentially single-valued* if when it selects more than one allocation, then all agents are indifferent between them: every agent gets the same utility from all tied allocations.

The following facts are well-known [see, e.g., Velez, 2018]. We include proofs for convenience.

Theorem 2.1. *(a) For every optimal room assignment σ, there are prices p so that (σ, p) is envy-free.*
(b) If (σ, p) is envy-free then σ is optimal.
(c) Let σ_1, σ_2 be optimal room assignments, and let (σ_1, p) be an envy-free allocation. Then (σ_2, p) is also an envy-free allocation, with all agents indifferent between the two: $v_{i\sigma_1(i)} - p_{\sigma_1(i)} = v_{i\sigma_2(i)} - p_{\sigma_2(i)}$ for all $i \in N$.

Proof. (a) An optimal room assignment σ forms a solution to the standard assignment problem [see, e.g., Wolsey, 1998, Section 4.3]. The dual of the assignment problem LP is

$$\min \sum_{i \in N} q_i + \sum_{r \in R} p_r \ \text{ s.t. } \ q_i + p_r \geqslant v_{ir} \ \text{ for } i \in N,\, r \in R.$$

Since σ is an optimal room assignment, by complementary slackness there exists a solution $(q_i), (p_r)$ to the dual program where $q_i + p_{\sigma(i)} = v_{i\sigma(i)}$ for each $i \in N$. Thus

$$v_{i\sigma(i)} - p_{\sigma(i)} = q_i \geqslant v_{ir} - p_r \quad \text{for all } i \in N,\, r \in R,$$

using dual feasibility. Thus (p_r) is an envy-free price vector, but we must ensure that $\sum_{r \in R} p_r = 1$, which we can do by adding a constant to each p_r. This preserves envy-freeness.

(b) Suppose (σ, p) is envy-free and σ' is any room assignment. Then $\sum_{i \in N} v_{i\sigma(i)} \geqslant \sum_{i \in N}(v_{i\sigma'(i)} - p_{\sigma'(i)} + p_{\sigma(i)}) = (\sum_{i \in N} v_{i\sigma'(i)}) - 1 + 1 = \sum_{i \in N} v_{i\sigma'(i)}$, where the inequality follows from envy-freeness. Thus σ has at least the welfare of σ'. Since σ' was arbitrary, σ is an optimal assignment.

(c) We show $v_{i\sigma_1(i)} - p_{\sigma_1(i)} = v_{i\sigma_2(i)} - p_{\sigma_2(i)}$ for $i \in N$. From this, envy-freeness of (σ_2, p) follows immediately. We have $v_{i\sigma_1(i)} - p_{\sigma_1(i)} \geqslant v_{i\sigma_2(i)} - p_{\sigma_2(i)}$ for all $i \in N$ since (σ_1, p) is envy-free. Sum these inequalities to get

$$(\textstyle\sum_{i \in N} v_{i\sigma_1(i)}) - 1 \geqslant (\sum_{i \in N} v_{i\sigma_2(i)}) - 1.$$

But the two sides of this inequality are equal, since both σ_1 and σ_2 are optimal. Hence each inequality is satisfied with equality, as required. □

Theorem 2.1(a) implies that an envy-free allocation exists for all valuation profiles. We can compute one in polynomial time: find an optimal room assignment σ using bipartite matching, then use linear programming to find prices p that make the allocation envy-free [Gal et al., 2017]. Theorem 2.1(c) implies that when selecting among envy-free allocations, we can restrict attention to any fixed σ and only vary the price vector p. By Theorem 2.1(c), all utility vectors achievable in an envy-free allocation are achieved by allocations of this form.

3. The Lexislack Solution

We start by considering a common form of robustness: we look for allocations that remain fair for all valuations that are within some radius of input valuations, for as large a radius as possible. Thus, unlike in later sections, we do not assume that valuations come from a probability distribution.

Let v be a valuation profile, fixed throughout. Let (σ, p) be an allocation. For $i \in N$ and $r \in R$, let

$$\Delta_{ir}(\sigma, p) = (v_{i\sigma(i)} - p_{\sigma(i)}) - (v_{ir} - p_r).$$

Then define the *slack* of this allocation as

$$\mathsf{slack}(\sigma, p) = \min_{i \in N} \min_{r \neq \sigma(i)} \Delta_{ir}(\sigma, p).$$

Thus, an allocation has positive slack if every agent strictly prefers their allocation to all other agents' allocations. An allocation (σ, p) is envy-free if and only if $\mathsf{slack}(\sigma, p) \geqslant 0$.

Slack is a measure of how robustly fair an allocation is, which we formalize in the following result.

Proposition 3.1. *Let (σ, p) be an envy-free allocation with $\mathsf{slack}(\sigma, p) = s \geqslant 0$. If v' is a valuation profile that is s-close to v in the sense that*

$$\|v_i - v'_i\|_1 = \textstyle\sum_{r \in R} |v_{ir} - v'_{ir}| \leqslant s$$

for all $i \in N$, then (σ, p) is also envy-free under v'.

Proof. Let $i, j \in N$. Then $\sum_{r \in R} |v_{ir} - v'_{ir}| \leqslant s$ implies

$$(v_{i\sigma(i)} - v'_{i\sigma(i)}) + (v'_{i\sigma(j)} - v_{i\sigma(j)}) \leqslant s \tag{1}$$

Adding $p_{\sigma(j)} - p_{\sigma(i)}$ to both sides and rearranging, we get

$$(v'_{i\sigma(i)} - p_{\sigma(i)}) - (v'_{i\sigma(j)} - p_{\sigma(j)}) \geqslant (v_{i\sigma(i)} - p_{\sigma(i)}) - (v_{i\sigma(j)} - p_{\sigma(j)}) - s \geqslant 0$$

where the last inequality is by definition of slack. Thus, i does not envy j under v'. Since i and j were arbitrary, it follows that (σ, p) is envy-free under v'. □

One can also prove variants of Proposition 3.1. For example, $\|v_i - v'_i\|_\infty \leqslant s/2$ also implies (1).[1]

If we wish to ensure robustness in a sense like in Proposition 3.1, this suggests the following rule:

$$\mathsf{maxislack}(v) = \operatorname{argmax}_{(\sigma,p)} \mathsf{slack}(\sigma, p).$$

[1] In future work, it may be interesting to study rules that explicitly maximize robustness defined with respect to ℓ_∞-distance rather than ℓ_1.

This rule always selects an envy-free allocation: since envy-free allocations exist for every v, there exists an allocation with non-negative slack, and hence the maxislack solution also has non-negative slack. A maxislack solution can be found in polynomial time by computing an optimal assignment σ and then solving the following LP:

$$\begin{aligned} \text{maximize} \quad & L \\ \text{subject to} \quad & (v_{i\sigma(i)} - p_{\sigma(i)}) - (v_{i\sigma(j)} - p_{\sigma(j)}) \geqslant L \quad \forall i \neq j \\ & \textstyle\sum_{r \in R} p_r = 1 \\ & p_r \in \mathbb{R} \quad \forall r \in R \end{aligned}$$

However, there are a few drawbacks to the maxislack rule. First, the rule is not essentially single-valued: there may be several maxislack allocations which induce different utilities. This is unlike Spliddit's maximin rule which is essentially single-valued [Alkan et al., 1991]. Second, there may be maxislack allocations that do not maximize robustness for *all* agents. To see this, suppose that two agents i_1 and i_2 agree on the valuation of every room. Then in any envy-free allocation, the utility they assign to the two bundles allocated to them is equal. Hence the maximum slack attainable is 0, and so *every* envy-free allocation is maxislack. However, there may be allocations for which the slack between other pairs of agents is larger than 0, and such allocations are more robustly fair.

In this spirit, to obtain robustness for a larger collection of agents (or of agent pairs), we can refine the maxislack solution using a leximin strategy. We call the resulting solution the *lexislack rule*. The lexislack rule selects an allocation (σ, p) that maximizes the smallest of the n^2 values $(\Delta_{ir}(\sigma, p))_{i \in N, r \in R}$, and subject to that maximizes the second-smallest of these values, and so on.

In contrast to the maxislack rule, the lexislack rule is essentially single-valued.

Theorem 3.2. *The lexislack rule is essentially single-valued.*

Proof. For now, fix an optimal room assignment σ. We show that there is a unique price vector p such that (σ, p) is a lexislack solution. Because the leximin relation over vectors is strictly convex, there is a unique vector $\Delta = \Delta(\sigma, p)$ maximizing the leximin objective, since if there were two different ones, a convex combination of the two would be strictly better. But Δ uniquely specifies a price vector: Δ gives the differences $p_r - p_{r'}$ between any pair of prices, and with $\sum_r p_r = 1$ this gives a unique price vector.

Next, we show that if σ_1 and σ_2 are optimal room assignments, and (σ_1, p) is an envy-free allocation, then $\Delta(\sigma_1, p) = \Delta(\sigma_2, p)$. By Theorem 2.1(c), (σ_2, p) is an allocation where every agent obtains the same utility as under (σ_1, p). Let $i \in N$. If $\sigma_1(i) = \sigma_2(i)$, then clearly the values Δ_{ir} are the same in both allocations. If $r_1 = \sigma_1(i) \neq \sigma_2(i) = r_2$, then equal utility under both allocations implies $v_{ir_1} - p_{r_1} = v_{ir_2} - p_{r_2}$, and hence $\Delta_{ir_2}(\sigma_1, p) = 0$ and $\Delta_{ir_1}(\sigma_2, p) = 0$. By definition, also $\Delta_{ir_1}(\sigma_1, p) = \Delta_{ir_2}(\sigma_2, p) = 0$, so the values of Δ_{ir_1} and Δ_{ir_2} agree on both allocations. For $r \in R \setminus \{r_1, r_2\}$, we have that the value of Δ_{ir} agrees on both allocations by the equal utility property. Hence $\Delta(\sigma_1, p) = \Delta(\sigma_2, p)$.

Thus, any vector $\Delta \geqslant 0$ achievable on one optimal room assignment can be achieved on any other optimal room assignment, with the same utility vector. This holds in particular for the lexislack vector Δ. We have seen that for any fixed room assignment, there is a unique lexislack utility vector. Hence the lexislack utility vector is unique. □

In addition, this rule remains efficiently computable.

Theorem 3.3. *A lexislack allocation can be found in polynomial time by solving $O(n^4)$ linear programs.*

Proof sketch. This can be done using standard techniques [see Kurokawa et al., 2018, Section 5]. We give an overview of the algorithm. Start by computing an optimal σ. We will decide on the best value of Δ_{ir} one-by-one. Let $F \leftarrow \emptyset$ be the set of (i, r) pairs for which we have fixed their value. Use linear programming to find a price vector such that (σ, p) maximizes the smallest of the non-fixed values Δ_{ir}, subject to keeping the other Δ_{ir} at their fixed value. Say the optimum is L. Now we need to find a pair $(i, r) \notin F$ such that necessarily $\Delta_{ir} = L$ in any lexislack allocation. This can again be done by linear programs that check whether it is possible that $\Delta_{ir} > L$. One can show that at least one such pair $(i, r) \notin F$ must exist; we then add it to F and fix its value to L, and repeat. □

4. Maximizing Probability of Envy-Freeness

In the previous section, we defined robustness using a measure of closeness based on the ℓ_1-distance. We now look at a more flexible model where true valuations are assumed to be noisy perturbations of the reported ones. A distribution $\mathcal{D}$ over valuations v, therefore, is obtained by asking agents for valuations, and then adding noise (e.g., Gaussian or uniform) around those valuations. Our goal will be to find an allocation (σ, p) that maximizes the probability of being envy-free with respect to $\mathcal{D}$, i.e., one that maximizes

$$\mathsf{EFrate}_{\mathcal{D}}(\sigma, p) = \Pr_{v\sim\mathcal{D}}[(\sigma, p) \text{ is envy-free under } v].$$

Our algorithmic approach for finding an allocation with high probability of envy-freeness is to obtain a sample S of m valuation profiles sampled from $\mathcal{D}$, and to compute an allocation that is envy-free on the most profiles in S, i.e., one that maximizes

$$\mathsf{EFrate}_{S}(\sigma, p) = \tfrac{1}{m} \cdot |\{v \in S : (\sigma, p) \text{ is envy-free under } v\}|.$$

If the number m of samples is sufficiently high, we may hope that the best allocation on the sample S is also approximately the best on the distribution $\mathcal{D}$. In this section, we will give a bound for the sample size m to be sufficient to ensure this property, and then we will discuss the computational problem of finding the best allocation for a given sample.

4.1. Sample Complexity

In this section, we will give an upper bound on the number of samples required to guarantee that the allocation that maximizes EFrate_S also (almost) maximizes $\mathsf{EFrate}_{\mathcal{D}}$, with high probability.

Theorem 4.1. *Let $\varepsilon, \delta > 0$. There is a value $m \in \mathbb{N}$ with*

$$m = O\left(\frac{n^2 \log n + \log(1/\delta)}{\varepsilon^2}\right)$$

such that for every probability distribution $\mathcal{D}$ over valuation profiles, if S is a collection of at least m samples drawn i.i.d. from $\mathcal{D}$, and (σ^*, p^*) is the allocation that maximizes EFrate_S, then with probability at least $1 - \delta$,

$$\mathsf{EFrate}_{\mathcal{D}}(\sigma^*, p^*) \geqslant \max_{(\sigma,p)} \mathsf{EFrate}_{\mathcal{D}}(\sigma, p) - \varepsilon.$$

We prove this theorem by adapting standard tools from learning theory. Let X be any set, with an unknown ground truth labeling $\tau : X \to \{0, 1\}$. A *hypothesis* is a function $h : X \to \{0, 1\}$. Given a *sample* $S = (x_1, \dots, x_m)$ of m elements of X (not necessarily distinct), write $\mathsf{err}_S(h) = \frac{1}{m}|\{x_i : h(x_i) \neq \tau(x_i)\}|$ for the fraction of samples that h labeled incorrectly. For a probability distribution $\mathcal{D}$ over X, write $\mathsf{err}_{\mathcal{D}}(h) = \Pr_{x\sim D}[h(x) \neq \tau(x)]$ for the probability that h incorrectly labels a point sampled from $\mathcal{D}$.

A *hypothesis class* $\mathcal{H}$ is a set of hypotheses. Given a random sample S drawn i.i.d. from $\mathcal{D}$, and knowledge of the true labeling τ of those samples, our goal is to find a hypothesis $h \in H$ that approximately minimizes $\mathsf{err}_{\mathcal{D}}(h)$, with high probability. Note that the ground truth τ, interpreted as a hypothesis, need not be a member of $\mathcal{H}$. In learning theory, this setup corresponds to "agnostic PAC learning", where the "realizability assumption" is not required to hold [Shalev-Shwartz and Ben-David, 2014, Section 3.2].

We say that a set $C \subseteq X$ is *shattered* by $\mathcal{H}$ if for all $S \subseteq C$, there exists $h \in \mathcal{H}$ with $h(x) = 1$ if $x \in S$ and $h(x) = 0$ if $x \in C \setminus S$. In other words, if we restrict the hypotheses in $\mathcal{H}$ to the set C, then all possible labelings of C are part of $\mathcal{H}$. The *VC dimension* $\mathsf{VCdim}(\mathcal{H})$ of $\mathcal{H}$ is the cardinality of the largest subset of X that is shattered by $\mathcal{H}$. We are interested in VC dimension due to the following standard result, adapted from Shalev-Shwartz and Ben-David [2014, Theorem 6.8], which says that PAC learning is possible on hypothesis classes of finite VC dimension.

Theorem 4.2. *Let $\varepsilon, \delta > 0$. Let $\mathcal{H}$ be a hypothesis class with $\mathsf{VCdim}(\mathcal{H}) = d$. Then there exists a value $m \in \mathbb{N}$ with*

$$m = O\left(\frac{d + \log(1/\delta)}{\varepsilon^2}\right)$$

such that for every probability distribution $\mathcal{D}$ over X, if S is a collection of at least m samples drawn i.i.d. from $\mathcal{D}$, and $h^* \in H$ is the hypothesis that minimizes err_S, then with probability at least $1 - \delta$,

$$\mathsf{err}_{\mathcal{D}}(h^*) \leqslant \min_{h\in\mathcal{H}} \mathsf{err}_{\mathcal{D}}(h) + \varepsilon.$$

For our application, we let X be the set of all valuation profiles v. The "correct" labeling is $\tau(v) = 1$ for all v. We identify allocations with hypotheses: For an allocation (σ, p), we define the hypothesis $h_{(\sigma,p)}$ so that for each v,

$$h_{(\sigma,p)}(v) = \begin{cases} 1 & \text{if } (\sigma, p) \text{ is envy-free under } v, \\ 0 & \text{otherwise.} \end{cases}$$

By these definitions, we have that for all S and $\mathcal{D}$,

$$\mathsf{EFrate}_S(\sigma, p) = 1 - \mathsf{err}_S(h_{(\sigma,p)}), \text{ and}$$
$$\mathsf{EFrate}_{\mathcal{D}}(\sigma, p) = 1 - \mathsf{err}_{\mathcal{D}}(h_{(\sigma,p)}).$$

We study the hypothesis class $\mathcal{H}$ of all such hypotheses:

$$\mathcal{H} = \{h_{(\sigma,p)} : \text{allocations } (\sigma, p)\}.$$

To bound its VC dimension, the following result is useful:

Lemma 4.3 (Shalev-Shwartz and Ben-David, 2014, Exercise 6.11)**.** *Let $\mathcal{H}_1, \ldots, \mathcal{H}_t$ be hypothesis classes over X, with $\mathsf{VCdim}(\mathcal{H}_i) \leqslant d$ for each $i = 1, \ldots, t$. Then*

$$\mathsf{VCdim}(\mathcal{H}_1 \cup \cdots \cup \mathcal{H}_t) \leqslant 4d \log(2d) + 2\log(t).$$

We can now bound the VC dimension of $\mathcal{H}$.

Lemma 4.4. $\mathsf{VCdim}(\mathcal{H}) = O(n^2 \log n)$.

Proof. For each room assignment σ, define the hypothesis class $\mathcal{H}_\sigma = \{h_{(\sigma,p)} : p \in \mathbb{R}^n\}$ corresponding to allocations whose room assignment is σ. Then $\mathcal{H} = \bigcup_\sigma \mathcal{H}_\sigma$ where the union ranges over all room assignments. We will show that $\mathsf{VCdim}(\mathcal{H}_\sigma) \leqslant n^2$ for each σ. Since there are $n!$ different room assignments and $\log n! = O(n \log n)$, it follows from Lemma 4.3 that $\mathsf{VCdim}(\mathcal{H}) = O(n^2 \log n)$, as required.

Let σ be a room assignment. Without loss of generality assume that $\sigma(i) = i$. Let $d \geqslant n^2 + 1$. Consider a collection of d distinct valuation profiles $v^{(1)}, \ldots, v^{(d)}$. We show that this collection cannot be shattered by $\mathcal{H}_\sigma$.

For $i, j \in N$, say $v^{(k)}$ is *uniquely restricting for* (i, j) if

$$v_{ij}^{(k)} - v_{ii}^{(k)} > v_{ij}^{(\ell)} - v_{ii}^{(\ell)} \quad \text{for all } \ell \neq k.$$

Thus, such a profile uniquely maximizes the amount by which agent i prefers j's room to her own room, ignoring prices. Clearly, for any pair $i, j \in N$, at most one profile can be uniquely restricting for it. Since there are n^2 many pairs (i, j) and $d > n^2$, there is at least one profile which is not uniquely restricting for any pair, say $v^{(1)}$.

We now ask if there is an allocation (σ, p) that is envy-free under $v^{(2)}, \ldots, v^{(d)}$, but not envy-free under $v^{(1)}$. We show that the answer is no, so $\mathcal{H}_\sigma$ fails to shatter this collection.

Assume for a contradiction that (σ, p) is such an allocation. Since it is not envy-free under $v^{(1)}$, there is a pair $i, j \in N$ with $v_{ij}^{(1)} - p_j > v_{ii}^{(1)} - p_i$ or equivalently

$$v_{ij}^{(1)} - v_{ii}^{(1)} > p_j - p_i. \tag{2}$$

As $v^{(1)}$ is not uniquely restricting for (i, j), for some $\ell \neq 1$,

$$v_{ij}^{(\ell)} - v_{ii}^{(\ell)} \geqslant v_{ij}^{(1)} - v_{ii}^{(1)}. \tag{3}$$

Combining (2) and (3), it follows that $v_{ij}^{(\ell)} - v_{ii}^{(\ell)} > p_j - p_i$. Thus, (σ, p) is not envy-free under $v^{(\ell)}$, a contradiction. □

Our main result in this section, Theorem 4.1, now follows immediately from Theorem 4.2.

4.2. Computational Complexity

To make use of Theorem 4.1, we need an algorithm that, given a collection $S = (v^{(1)}, \dots, v^{(m)})$ of valuation profiles sampled from $\mathcal{D}$, finds an allocation that maximizes $\mathsf{EFrate}_S(\sigma, p)$. This problem can be encoded as an integer linear program via standard encoding techniques, using binary variables x_{ir} encoding that agent i receives room r, continuous variables p_r encoding the prices, and a binary variable y_ℓ for each sample $\ell \in [m]$, indicating whether the produced allocation will be envy-free under $v^{(\ell)}$. The full encoding appears in Appendix B.

Instead of an ILP approach, can we hope for a polynomial time algorithm finding the best allocation? Let us formulate our optimization problem as a decision problem as follows.

> EF-Rate Maximization
> **Input:** Set N of agents, set R of rooms, a list of m valuation profiles $v^{(1)}, \dots, v^{(m)}$, number B.
> **Question**: Does there exist an allocation that is envy-free for at least B of the m valuation profiles?

Unfortunately, this problem is computationally hard.

Theorem 4.5. EF-Rate Maximization *is NP-complete, even for binary valuation profiles ($v_{ir} \in \{0, 1\}$).*

Proof. Membership in NP is clear. We give a reduction from Clique. Let $G = (V, E)$ be a graph with n vertices and m edges and let k be the target clique size.

We make each vertex an agent, $N = V$. The set of rooms is $R = \{r_1, \dots, r_k, d_1, \dots, d_{n-k}\}$ consisting of k *slot rooms* and of $n - k$ *dummy rooms*. Writing $E = \{e_1, \dots, e_m\}$, we construct m valuation profiles, one per edge. For $\ell \in [m]$, write $e_\ell = \{u, v\}$; the valuation profile $v^{(\ell)}$ is defined by

$$v^{(\ell)}_{i,r} = \begin{cases} 1 & \text{if } i \in \{u, v\} \text{ and } r \in \{r_1, \dots, r_k\}, \\ 0 & \text{otherwise.} \end{cases}$$

Thus, in the ℓth valuation profile, the two agents corresponding to the endpoints of the ℓth edge want to be in a slot room. All other agents do not care. Finally, set $B = \binom{k}{2}$.

We prove that G has a k-clique iff there is an allocation that is envy-free in at least B of the valuation profiles.

($\Leftarrow$): Suppose (σ, p) is envy-free for B profiles. Let $C \subseteq V$ be the set of k agents/vertices that are assigned to slot rooms under σ; write $C = \{i_1, \dots, i_k\}$. Let $\ell_1, \dots, \ell_B$ be the collection of indices corresponding to valuation profiles under which the allocation is envy-free. We claim that C is a clique, by showing that $e_{\ell_t} \subseteq C$ for each $t \in [B]$. This suffices since a set of k vertices with $\binom{k}{2}$ edges is a clique.

Let $t \in [B]$. Since (σ, p) is envy-free under $v^{(\ell_t)}$, by Theorem 2.1(b), σ is optimal under $v^{(\ell_t)}$. This implies that σ has welfare 2, which happens only if both endpoints of edge e_{ℓ_t} get a slot room. So by definition of C, $e_{\ell_t} \subseteq C$, as desired.

($\Rightarrow$): Suppose there is a clique $C \subseteq V$ of size k in G; write $C = \{i_1, \dots, i_k\}$. Make a room assignment σ in which we assign agent i_s to slot room r_s, for each $s \in [k]$. The remaining agents can be assigned arbitrarily to dummy rooms. We set $p_r = \frac{1}{n}$ for each $r \in R$.

Write $e_{\ell_1}, \dots, e_{\ell_B}$ for the set of edges within C; there are exactly B of them since C is a clique. Let $t \in [B]$, and write $e_{\ell_t} = \{i_a, i_b\}$. We show that (σ, p) is envy-free under $v^{(\ell_t)}$. All agents except i_a and i_b are indifferent between all rooms, and since all rents are the same, they are not envious. Agents i_a and i_b both receive a room that they most prefer, and since all rents are the same, they are not envious. □

There are two sources of computational difficulty for solving EF-Rate Maximization: we have to decide on one of the $n!$ possible room assignments, and we have to decide on which subset of valuation profiles we are aiming to be envy-free on. But in practice, there is a way to avoid the first source of hardness. Suppose the m valuation profiles are sampled from a *continuous* distribution $\mathcal{D}$. Then with probability 1, for each sampled profile $v^{(\ell)}$ there is a *unique* optimal room assignment $\sigma^{(\ell)}$. Any solution to the EF-rate maximization problem must use a room assignment that is optimal for at least one of the given valuation profiles. Thus, at most m different room assignments are candidates, and we can find an optimal solution using m calls to the following problem (one call for each candidate assignment $\sigma^{(\ell)}$):

> EF-Rate Maximization with Fixed Assignment
> **Input:** A list of m valuation profiles $v^{(1)}, \dots, v^{(m)}$, number B, room assignment σ.
> **Question**: Is there a price vector p such that (σ, p) is envy-free for at least B of the m valuation profiles?

Unfortunately, this version of the problem is also hard, and so this trick for continuous distributions does not help. We prove this by reduction from the feedback arc set problem. The proof is in Appendix C.1.

Theorem 4.6. EF-Rate Maximization with Fixed Assignment *is NP-complete.*

Nevertheless, as we show in Section 6, we can solve this problem in practice using integer linear programming (ILP). The reason this is possible is that the sample complexity is relatively low, leading to an ILP of practical size.

5. Minimizing Expected Envy

In Section 4, we defined robust envy-freeness as allocations that have a high probability of being envy-free when valuations come from a given distribution $\mathcal{D}$. In this section, we consider a different objective function that is more fine-grained. In measuring the probability of envy-freeness, we implicitly treat all failures of envy-freeness equally. We will now minimize *expected envy*, which treats cases where one agent envies another by a lot as more severe.

Given a valuation profile v and an allocation (σ, p), we define the allocation's *(maximum) envy*, $\mathsf{envy}_v(\sigma, p)$, to be

$$\max\left\{0, \max_{i,j\in N} \left[(v_{i\sigma(j)} - p_{\sigma(j)}) - (v_{i\sigma(i)} - p_{\sigma(i)})\right]\right\}.$$

This quantity, which is related to slack as considered in Section 3, measures the biggest amount by which one agent prefers another's bundle. In principle one could allow negative values of $\mathsf{envy}_v(\sigma, p)$ for allocations that have positive slack, but we chose to force these values to be non-negative, since our focus is on avoiding envy. Note that an allocation is envy-free if and only if $\mathsf{envy}_v(\sigma, p) = 0$.

Our goal in this section is to find an allocation minimizing the expected envy with respect to $\mathcal{D}$, defined as

$$\mathsf{envy}_{\mathcal{D}}(\sigma, p) = \mathbb{E}_{v\sim\mathcal{D}}[\mathsf{envy}_v(\sigma, p)].$$

Our approach will be similar to before: we obtain a sufficiently large sample S of m profiles from $\mathcal{D}$ and select the allocation that does best on the sample, i.e. it minimizes

$$\mathsf{envy}_S(\sigma, p) = \tfrac{1}{m}\textstyle\sum_{v\in S} \mathsf{envy}_v(\sigma, p).$$

5.1. Sample Complexity

For stating our sample complexity bound, we assume that valuations v are normalized: let $v_{ir} \geqslant 0$ for all $i \in N$ and $r \in R$, and $\sum_{r\in R} v_{ir} = 1$ for all $i \in N$. We are going to prove the following result:

Theorem 5.1. *Let $\varepsilon, \delta > 0$, and let $\mathcal{D}$ be a distribution. If we draw $m = O(\frac{n}{\varepsilon^2}\log\frac{n}{\varepsilon\delta})$ samples i.i.d. from $\mathcal{D}$ and if (σ^*, p^*) minimizes envy_S , then with probability at least $1-\delta$, we have*

$$\mathsf{envy}_{\mathcal{D}}(\sigma^*, p^*) < \min_{(\sigma,p)} \mathsf{envy}_{\mathcal{D}}(\sigma, p) + \varepsilon.$$

Thus, if we draw sufficiently many samples, then with high probability the allocation minimizing expected envy on the sample will, up to ε, be minimizing with respect to $\mathcal{D}$.

We prove this result by discretizing the space of allocations. We then use a concentration inequality to show that w.h.p. the expected envy with respect to $\mathcal{D}$ is close to the expected envy with respect to the sample S.

We start by proving a few technical lemmas. First, define Λ to be the set all allocations (σ, p) with $-2 \leqslant p_r \leqslant 2$ for all $r \in R$. We call such allocations *reasonable.* Our first lemma shows that we may restrict attention to reasonable allocations only: in particular, if an allocation minimizes $\mathsf{envy}_{\mathcal{D}}$ then it must be reasonable.

Lemma 5.2. *Let σ be a room assignment and v a profile.*

(a) If p is a price vector with $|p_r - p'_r| > 2$ for some $r, r' \in R$, then $\mathsf{envy}_v(\sigma, p) > 1$.

(b) If $p = (\frac{1}{n}, \ldots, \frac{1}{n})$, then $\mathsf{envy}_v(\sigma, p) \leqslant 1$.

(c) If (σ, p) is reasonable, then $\mathsf{envy}_v(\sigma, p) \leqslant 5$.

Proof. Note that since valuations are assumed to sum to 1, we have $v_{ir} - v_{ir'} \leqslant 1$ for all $r, r' \in R$.

(a) Say $p_r > p_{r'} + 2$ and $\sigma(i) = r$. Since $v_{ir} \leqslant v_{ir'} + 1$ as just noted, we have $v_{ir} - p_r < v_{ir'} - p_{r'} - 1$.

(b) If prices for all rooms are equal, then $\mathsf{envy}_v(\sigma, p) = \max\{0, \max_{i,j}(v_{i\sigma(j)} - v_{i\sigma(i)})\} \leqslant \max\{0, 1\} = 1$.

(c) $|v_{i\sigma(j)} - v_{i\sigma(i)}| + |p_{\sigma(j)} - p_{\sigma(i)}| \leqslant 1 + 4 = 5$. □

From now on, we assume all allocations to be reasonable.

Our second lemma says that if two allocations have similar price vectors, then they have similar expected envy.

Lemma 5.3. *Let $p, p' \in \mathbb{R}^n$ be such that $|p_r - p'_r| \leqslant t$ for all $r \in R$. Then for any sample S and distribution $\mathcal{D}$,*

$$|\mathsf{envy}_{\mathcal{D}}(\sigma, p) - \mathsf{envy}_{\mathcal{D}}(\sigma, p')| \leqslant 2t,$$
$$|\mathsf{envy}_{S}(\sigma, p) - \mathsf{envy}_{S}(\sigma, p')| \leqslant 2t.$$

Proof. First, we claim that for any valuation profile v,

$$|\mathsf{envy}_{v}(\sigma, p) - \mathsf{envy}_{v}(\sigma, p')| \leqslant 2t.$$

This holds since the value of $(v_{i\sigma(j)} - p_{\sigma(j)}) - (v_{i\sigma(i)} - p_{\sigma(i)})$ changes by at most $\pm 2t$ if we move from p to p', and thus the same holds after taking the maximum.

Now let $\mathcal{D}$ be a distribution. By linearity of expectation, and since $|\mathbb{E}[X]| \leqslant \mathbb{E}[|X|]$ by Jensen's inequality,

$$\begin{aligned} &|\mathsf{envy}_{\mathcal{D}}(\sigma, p) - \mathsf{envy}_{\mathcal{D}}(\sigma, p')| \\ &= |\mathbb{E}_{v\sim\mathcal{D}}[\mathsf{envy}_{v}(\sigma, p) - \mathsf{envy}_{v}(\sigma, p')]| \\ &\leqslant \mathbb{E}_{v\sim\mathcal{D}}\left[|\mathsf{envy}_{v}(\sigma, p) - \mathsf{envy}_{v}(\sigma, p')|\right] \leqslant 2t, \end{aligned}$$

where the last inequality follows by our claim. This proves the first statement. The second statement follows from the first by taking $\mathcal{D}$ to be the uniform distribution over S. □

We also need a standard concentration inequality.

Lemma 5.4 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_m$ be i.i.d. random variables with $0 \leqslant X_k \leqslant c$ and $\mathbb{E}[X_k] = \mu$ for all $k \in [m]$. Then for all $\varepsilon > 0$,*

$$\Pr\left[|\mu - \tfrac{1}{m}\textstyle\sum_{k=1}^{m} X_i| \geqslant \varepsilon\right] \leqslant 2\exp(-2m\varepsilon^2/c^2).$$

We are now ready to prove our main result of this section.

Proof of Theorem 5.1. Let $t = 1/\lceil 12/\varepsilon \rceil$. Let $\Lambda^t \subseteq \Lambda$ be the set of all *discretized allocations* (σ, p) where p_r is an integer multiple of t. Note that for any $(\sigma, p) \in \Lambda$, there is a discretized allocation $(\sigma, p') \in \Lambda^t$ with $|p_r - p'_r| \leqslant t$ for all r (call such allocations *t-close*), obtained by rounding the values p_r up or down to ensure that $\sum_r p'_r = 1$.

Let S be a random sample from $\mathcal{D}$ of size m, where

$$m = \tfrac{200}{\varepsilon^2} \ln\left[\left(\tfrac{60}{\varepsilon}\right)^n \tfrac{2n!}{\delta}\right] = O\left(\tfrac{n}{\varepsilon^2} \ln\left(\tfrac{n}{\varepsilon\delta}\right)\right).$$

Now write:

- $\mathrm{OPT}_{\mathcal{D}}$ for an allocation (σ, p) minimizing $\mathsf{envy}_{\mathcal{D}}$,
- OPT_{S} for an allocation minimizing envy_{S} (which depends on the random choice of S),
- $\overline{\mathrm{OPT}}_{\mathcal{D}} \in \Lambda^t$ for a discretized allocation t-close to $\mathrm{OPT}_{\mathcal{D}}$,
- $\overline{\mathrm{OPT}}_{S} \in \Lambda^t$ for a discretized allocation t-close to OPT_{S}.

Let $(\sigma, p) \in \Lambda$. For $k \in [m]$, let X_k be the random variable taking the value $\mathsf{envy}_v(\sigma, p)$, where v is the kth sample in S. By reasonableness and Lemma 5.2, $0 \leqslant X_k \leqslant 5$. Then Hoeffding's inequality implies that

$$\Pr\left[|\mathsf{envy}_{S}(\sigma, p) - \mathsf{envy}_{\mathcal{D}}(\sigma, p)| \geqslant \tfrac{\varepsilon}{4}\right] \leqslant 2\exp(-\tfrac{2}{25}m(\tfrac{\varepsilon}{4})^2).$$

Let E be the event that $|\mathsf{envy}_S(\sigma, p) - \mathsf{envy}_\mathcal{D}(\sigma, p)| < \varepsilon/4$ holds for all discretized allocations $(\sigma, p) \in \Lambda^t$ simultaneously. By Hoeffding's inequality and a union bound over all $|\Lambda^t| \leqslant (\frac{4}{t})^n n!$ discretized allocations, we get

$$\Pr[E] \geqslant 1 - (\tfrac{4}{t})^n n!\, 2\exp(-\tfrac{2}{25} m (\tfrac{\varepsilon}{4})^2) \geqslant 1 - \delta.$$

where the second inequality holds by choice of m.

Suppose that the event E attains. In this case we have:

$$\begin{aligned}
&\mathsf{envy}_\mathcal{D}(\textsc{opt}_S) - \mathsf{envy}_\mathcal{D}(\textsc{opt}_\mathcal{D}) \\
&= (\mathsf{envy}_\mathcal{D}(\textsc{opt}_S) - \mathsf{envy}_\mathcal{D}(\overline{\textsc{opt}}_S)) && \text{(Lemma 5.3)} \\
&\quad + (\mathsf{envy}_\mathcal{D}(\overline{\textsc{opt}}_S) - \mathsf{envy}_S(\overline{\textsc{opt}}_S)) && \text{(E attains)} \\
&\quad + (\mathsf{envy}_S(\overline{\textsc{opt}}_S) - \mathsf{envy}_S(\textsc{opt}_S)) && \text{(Lemma 5.3)} \\
&\quad + (\mathsf{envy}_S(\textsc{opt}_S) - \mathsf{envy}_S(\overline{\textsc{opt}}_\mathcal{D})) && \text{(optimality)} \\
&\quad + (\mathsf{envy}_S(\overline{\textsc{opt}}_\mathcal{D}) - \mathsf{envy}_\mathcal{D}(\overline{\textsc{opt}}_\mathcal{D})) && \text{(E attains)} \\
&\quad + (\mathsf{envy}_\mathcal{D}(\overline{\textsc{opt}}_\mathcal{D}) - \mathsf{envy}_\mathcal{D}(\textsc{opt}_\mathcal{D})) && \text{(Lemma 5.3)} \\
&< 2t + \varepsilon/4 + 2t + 0 + \varepsilon/4 + 2t \\
&= 6/\lceil 12/\varepsilon \rceil + \varepsilon/2 \leqslant \varepsilon.
\end{aligned}$$

The references on the right indicate what we have used to bound the respective term to obtain the strict inequality. "Optimality" refers to the fact that $\textsc{opt}_S$ minimizes envy_S. Because event E implies the above inequality, we see that

$$\Pr[\mathsf{envy}_\mathcal{D}(\textsc{opt}_S) - \mathsf{envy}_\mathcal{D}(\textsc{opt}_\mathcal{D}) < \varepsilon] \geqslant \Pr[E] \geqslant 1 - \delta.$$

This proves the result. □

Note that for this result we employed a direct approach. This technique and its discretization step would not have worked for the envy-free rate, because two very close rent divisions can in principle have very different EF rates. On the other hand, we expect that similar bounds for the minimum envy objective could be obtained by using extensions of VC dimension to real-valued functions (e.g., pseudo-dimension).

5.2. Computational Complexity

Again, our sample complexity result needs an algorithm that finds the best allocation for a given sample S. Like for EFrate, we can solve this problem using integer linear programming (see Appendix B). For the formal complexity analysis, consider the following decision problem:

Expected Envy Minimization **Input:** List $S = (v^{(1)}, \ldots, v^{(m)})$, number B. **Question**: Is there (σ, p) with $\mathsf{envy}_S(\sigma, p) \leqslant B$?

This problem is again NP-complete. The proof is in Appendix C.2 and uses a similar reduction from Clique as before.

Theorem 5.5. Expected Envy Minimization *is NP-complete, even for binary valuation profiles.*

Interestingly, this problem becomes easy once we fix a room assignment σ, because the best price vector can then be computed by linear programming (because the values of the integer variables in the ILP shown in Appendix B are decided by the fixed room assignment σ). In particular, this means that the problem can be solved in time $n! \cdot \mathrm{poly}(n, m)$, and thus is fixed-parameter tractable with respect to the number of agents n. This is good news: instances will often have a small number of agents, but we will want to consider as large a sample as feasible to ensure low maximum envy. In fact, as we will see momentarily, the NP-completeness of the problem is not an obstacle in real-world instances.

6. Experiments

We evaluated our rules on user data taken from Spliddit.[2] We studied distributions obtained by adding noise to valuations. We started by selecting 1,000 instances v at random, to speed up computations. The

[2]This dataset was kindly provided to us in anonymized form by the maintainer of Spliddit, Nisarg Shah.

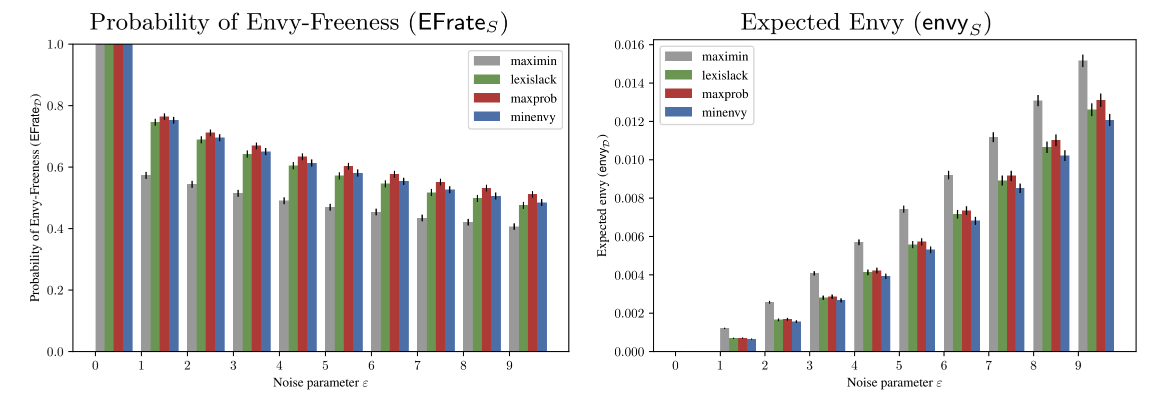


Figure 1: Results of experiments for the Uniform noise model.

same selection is used for each experiment. For each instance, we normalize the rent to 1, and normalize valuations to sum to 1. We considered three noise models, each parameterized by a choice of noise level $\varepsilon \in \{0, 0.01, \dots, 0.09\}$.

$$v_{ir}^{(\ell)} \sim v_{ir} \cdot (1 + \text{Uniform}[-\varepsilon, +\varepsilon]) \qquad \text{(Uniform)}$$
$$v_{ir}^{(\ell)} \sim v_{ir} \cdot (1 + \text{N}[0, \varepsilon]) \qquad \text{(Normal)}$$
$$v_{ir}^{(\ell)} \sim v_{ir} \cdot (1 + r \cdot \text{N}[0, \varepsilon]) \qquad \text{(Biased Normal)}$$

In each of these noise models, valuations are increased or decreased by a random fraction. Here, $N[\mu, \sigma]$ is a normal distribution with mean μ and standard deviation σ. For the biased normal noise model, we put rooms in an arbitrary fixed order and label them with integers $0, 1, \dots, n-1$. Rooms with a higher index have more noise.

For each noise model and choice of ε, we produced a sample S of size $m = 100$. We then computed allocations maximizing EFrate_S and minimizing envy_S. We also computed the maximin and lexislack rules based on the input profile v. For each of the four allocations, we calculated their value of EFrate_S, and of envy_S. We then average over all 1,000 instances. The results are shown in Figure 1 for the Uniform noise model. Results for the other noise models and more details are given in Appendix A. As expected, on each of the two metrics, the rule optimizing it does best, but all three rules aiming for robustness do similarly well. Spliddit's maximin rule does significantly worse on our metrics. Before the experiments, we expected that lexislack would do worse for the biased noise model, but this does not appear to be the case.

In the appendix, we also evaluate the sampling-based rules on a freshly drawn sample different from the sample used to optimize the rules (Appendix A.2) as well as on a sample drawn from a different probability distribution (Appendix A.3), to evaluate how sensitive these methods are to being optimized on a small sample and to knowing the 'right' noise distribution. In both cases, we find that the performance of the sampling-based methods worsens, while lexislack continues to be robust.

Figure 2 shows average computation time to compute allocations optimizing EFrate_S and envy_S, using Gurobi 9.1.2 on four threads of an AMD Ryzen 2990WX (128 GB RAM) with the ILP formulations from Appendix B. The results were obtained for a random selection of 300 Spliddit instances with $n = 4$, with the Uniform noise model for $\varepsilon = 0.05$, and sample sizes m varying from 1 to 125. Minimizing envy is much faster due to fewer integral variables. In Appendix A.4, we report some additional computation times as a function of the noise parameter ε.

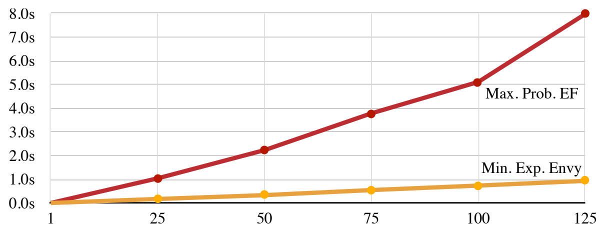


Figure 2: Computation time depending on sample size

7. Future Directions

Our approach should be applicable to many settings beyond rent division, such as homogeneous divisible goods, cake cutting, or even indivisible goods. For example, the lexislack rule can be adapted to these settings, and similar results as in our distribution-based approach might be achievable.

We have shown that the lexislack rule shares some key properties with the maximin rule, such as essential single-valuedness and polynomial-time computability. It would be interesting to axiomatically contrast the two solutions, for example with respect to strategic properties like manipulability.

For our distribution-based approach, we assumed that we have access to $\mathcal{D}$ only via sampling. Often we may know $\mathcal{D}$ more explicitly, for example if we are just adding noise to reported valuations. For such well-behaved $\mathcal{D}$, can we design direct algorithms for finding optimal allocations with respect to our two objectives, without needing samples?

Acknowledgments

This work was partially supported by the National Science Foundation under grants IIS-2147187, CCF-2007080, IIS-2024287, and CCF-1733556; and by the Office of Naval Research under grant N00014-20-1-2488.

References

A. Alkan, G. Demange, and D. Gale. Fair allocation of indivisible goods and criteria of justice. *Econometrica*, 59(4):1023–1039, 1991. [→ p. 3, 5]

E. Aragones. A derivation of the money Rawlsian solution. *Social Choice and Welfare*, 12:267–276, 1995. [→ p. 3]

H. Aziz, P. Biró, R. de Haan, and B. Rastegari. Pareto optimal allocation under uncertain preferences: uncertainty models, algorithms, and complexity. *Artificial Intelligence*, 276:57–78, 2019. [→ p. 3]

H. Aziz, P. Biró, S. Gaspers, R. de Haan, N. Mattei, and B. Rastegari. Stable matching with uncertain linear preferences. *Algorithmica*, 82(5):1410–1433, 2020. [→ p. 3]

M.-F. Balcan. Data-driven algorithm design. In T. Roughgarden, editor, *Beyond the Worst-Case Analysis of Algorithms*, chapter 29. Cambridge University Press, 2020. [→ p. 3]

X. Bei, Z. Li, J. Liu, S. Liu, and X. Lu. Fair division of mixed divisible and indivisible goods. *Artificial Intelligence*, 293:103436, 2021. [→ p. 3]

R. Bredereck, P. Faliszewski, A. Kaczmarczyk, R. Niedermeier, P. Skowron, and N. Talmon. Robustness among multiwinner voting rules. *Artificial Intelligence*, 290:Article 103403, 2021. [→ p. 3]

E. Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011. [→ p. 1]

I. Caragiannis, D. Kurokawa, H. Moulin, A. D. Procaccia, N. Shah, and J. Wang. The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation (TEAC)*, 7(3): 1–32, 2019. [→ p. 1]

J. Chen, P. Skowron, and M. Sorge. Matchings under preferences: Strength of stability and trade-offs. In *Proceedings of the 2019 ACM Conference on Economics and Computation (EC)*, pages 41–59, 2019. [→ p. 3]

A. Critch. Robust rental harmony. *https://acritch.com/rent/*, 2015. Archived at https://perma.cc/Q9MD-HMAY. [→ p. 3]

B. Flanigan, P. Gölz, A. Gupta, B. Hennig, and A. D. Procaccia. Fair algorithms for selecting citizens' assemblies. *Nature*, 596:548–552, 2021. [→ p. 1]

Y. Gal, M. Mash, A. D. Procaccia, and Y. Zick. Which is the fairest (rent division) of them all? *Journal of the ACM*, 64(6):Article 39, 2017. [→ p. 1, 3, 4]

G. Gawron and P. Faliszewski. Robustness of approval-based multiwinner voting rules. In *Proceedings of the 6th International Conference on Algorithmic Decision Theory (ADT)*, pages 17–31, 2019. [→ p. 3]

J. Goldman and A. D. Procaccia. Spliddit: Unleashing fair division algorithms. *SIGecom Exchanges*, 13 (2):41–46, 2014. [→ p. 1]

D. Kurokawa, A. D. Procaccia, and N. Shah. Leximin allocations in the real world. *ACM Transactions on Economics and Computation*, 6(3–4):Article 11, 2018. [→ p. 5]

T. Mai and V. V. Vazirani. Finding stable matchings that are robust to errors in the input. In *Proceedings of the 26th Annual European Symposium on Algorithms (ESA)*, page Article No. 60, 2018. [→ p. 3]

V. Menon and K. Larson. Algorithmic stability in fair allocation of indivisible goods among two agents. *arXiv:2007.15203*, 2020. [→ p. 3]

N. Misra and C. Sonar. Robustness radius for Chamberlin–Courant on restricted domains. In *Proceedings of the 45th International Conference on Current Trends in Theory and Practice of Informatics (SOFSEM)*, pages 341–353, 2019. [→ p. 3]

H. Moulin. Fair division in the internet age. *Annual Review of Economics*, 11:407–441, 2019. [→ p. 1]

A. D. Procaccia, R. A. Velez, and D. Yu. Fair rent division on a budget. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*, pages 1177–1184, 2018. [→ p. 3]

S. Shalev-Shwartz and S. Ben-David. *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014. [→ p. 6, 7]

D. Shiryaev, L. Yu, and E. Elkind. On elections with robust winners. In *Proceedings of the 12th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 415–422, 2013. [→ p. 3]

F. E. Su. Rental harmony: Sperner's lemma in fair division. *American Mathematical Monthly*, 106(10): 930–942, 1999. [→ p. 3]

L.-G. Svensson. Large indivisibles: An analysis with respect to price equilibrium and fairness. *Econometrica*, 51(4):939–954, 1983. [→ p. 1, 3]

R. A. Velez. Equitable rent division. *ACM Transactions on Economics and Computation (TEAC)*, 6(2): Article 9, 2018. [→ p. 3]

L. A. Wolsey. *Integer Programming*. John Wiley & Sons, 1998. [→ p. 3]

A. Experimental Results

In this section, we will present more detailed results of the experiments described in Section 6 of the main body of the paper. We describe the results in four subsections:

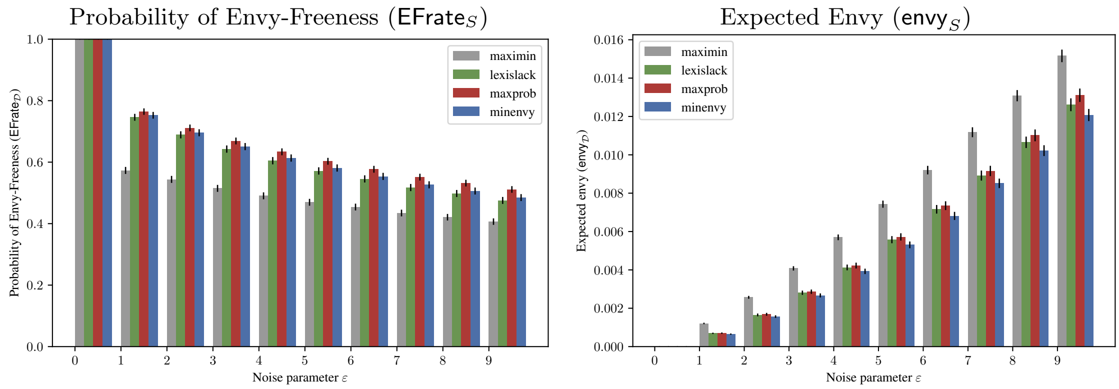
- In Appendix A.1, we evaluate the rules by calculating their performance in terms of envy-rate and expected envy on the same sample ($m = 100$) that we used to optimize the two probabilistic rules (the *optimization sample*).
- In Appendix A.2, we evaluate the rules by calculating their performance on a fresh sample ($m = 1000$) that is different from the one used to compute the rules.
- In Appendix A.3, we evaluate our probabilistic rules trained on one specific noise model (normal noise with $\varepsilon = 0.05$) on the other noise models.
- In Appendix A.4 we briefly discuss computation times.

A.1. Evaluation on the Optimization Sample

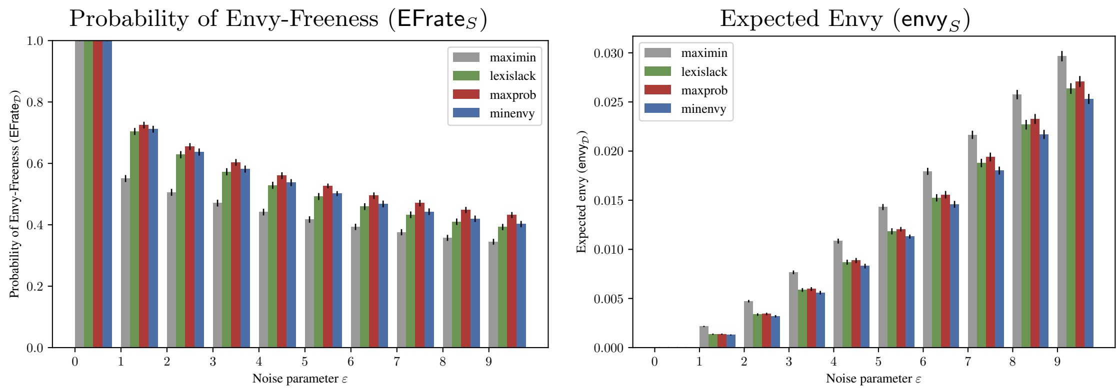
When evaluating the rules based on the optimization sample (or in other terminology, if we take the test set to be the same as the training set), then our optimizing rules are optimal by definition. Indeed, in each of the charts, we can see that the maxprob rule (maximizing the probability of envy-freeness) has the best performance out of all the rules with respect to the probability of envy-freeness; and analogously for the minenvy rule and the envy objective. When evaluating on the optimization sample, we see that the respective optimizing rule outperforms lexislack, but only by a modest amount.

In all the charts, error bars show standard errors, which are small due to the large number of instances.

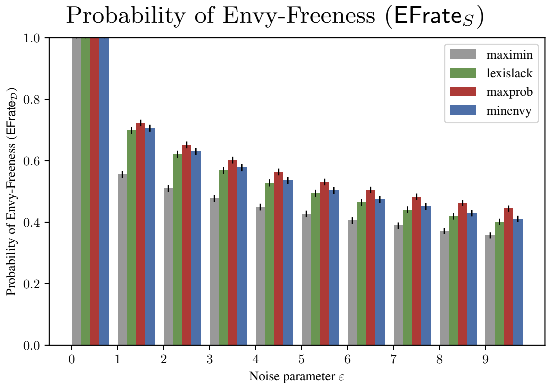
A.1.1. Uniform Noise

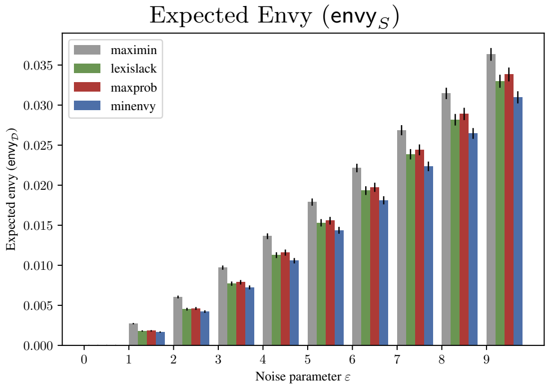


A.1.2. Normal Noise



A.1.3. Biased Normal Noise

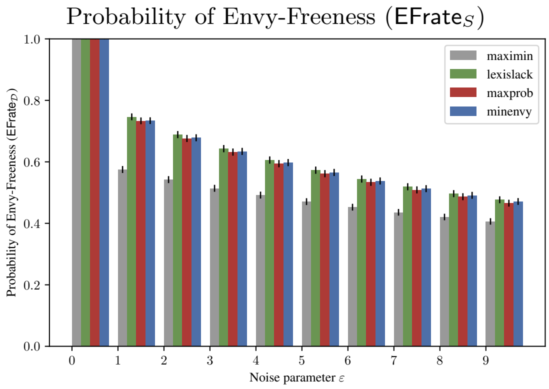


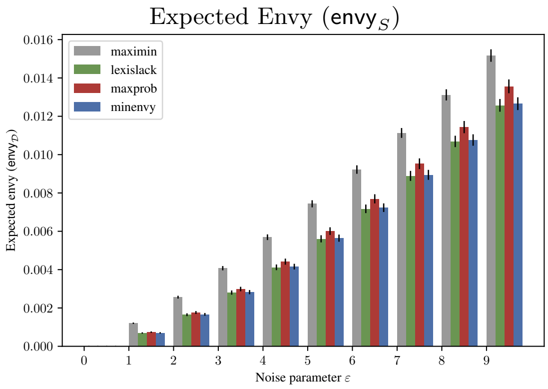


A.2. Evaluation on a Fresh Sample

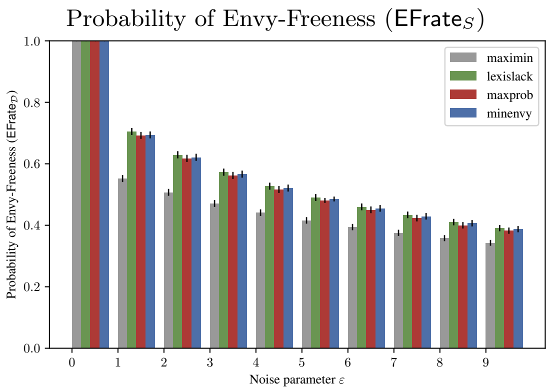
For these charts, we drew a fresh sample of size $m = 1000$, and calculated the values EFrate_S and envy_S with respect to this fresh sample. (Since evaluation is much cheaper than optimization, it is no problem to use a large sample size.) We see that the advantage of the optimizing rules over lexislack shrinks or disappears. The minenvy rule performs the same as lexislack for the uniform and normal noise models with respect to the envy objective, though it outperforms lexislack by an extremely small amount for the biased normal noise model. On the other hand, surprisingly, the maxprob rule does strictly worse than both the minenvy and lexislack rules, on the probability objective. This suggests that the sample size of $m = 100$ was too small to allow the maxprop rule to properly generalize to the underlying noise distribution.

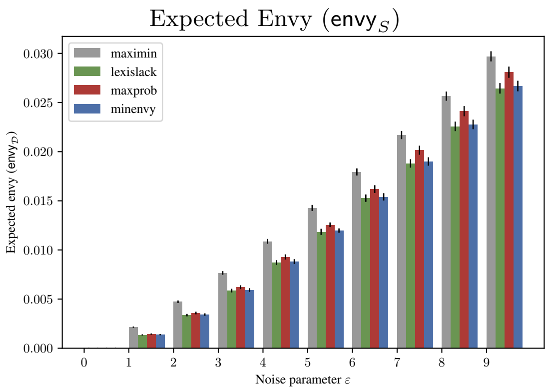
A.2.1. Uniform Noise



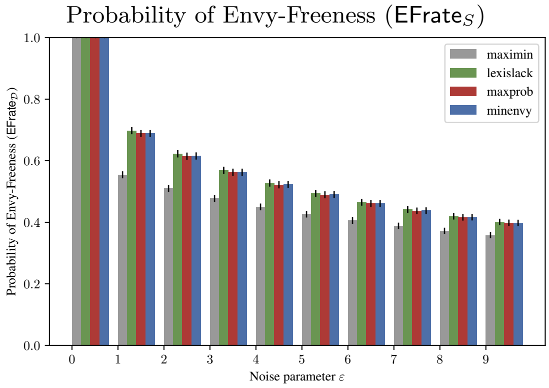


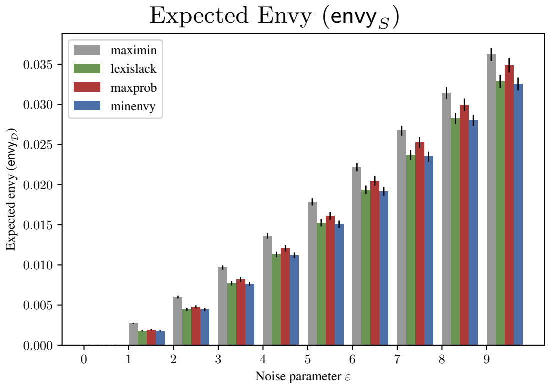
A.2.2. Normal Noise





A.2.3. Biased Normal Noise



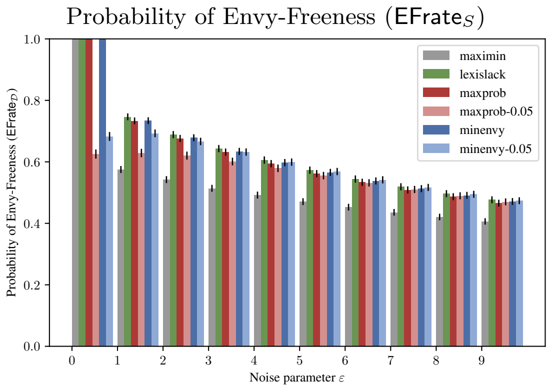


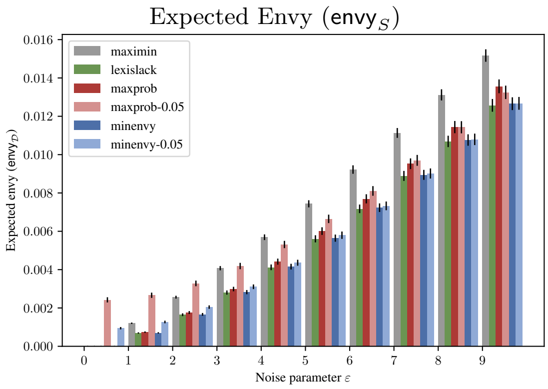
A.3. Evaluation of Rules Trained on a Different Distribution

Like in the previous subsection, the following charts are with respect to a fresh sample of size $m = 1000$. However, in each chart, we now add two 'new' rules, namely the allocations selected by the maxprob and minenvy rules when optimized on samples drawn from the normal noise model with $\varepsilon = 0.05$. Thus, these charts allow us to gauge the performance of the distribution-based methods when they are optimized using the 'wrong' distribution.

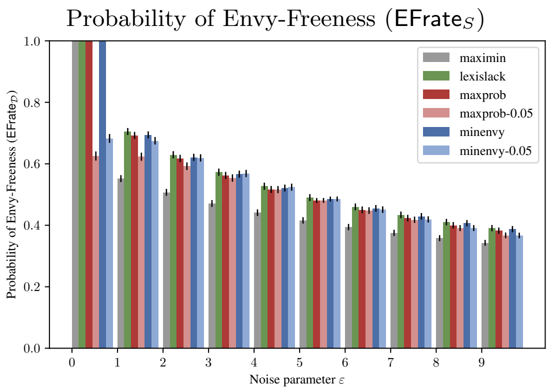
For the probability objective and the uniform and normal noise models, not much performance is lost. However, the rules optimized for the normal noise models perform poorly for the biased noise models (compared to the rules optimized for that model, and also compared to lexislack). For the envy objective, we see notably bad performance for uniform noise.

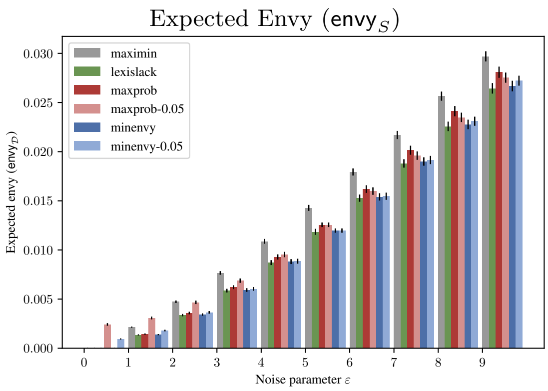
A.3.1. Uniform Noise



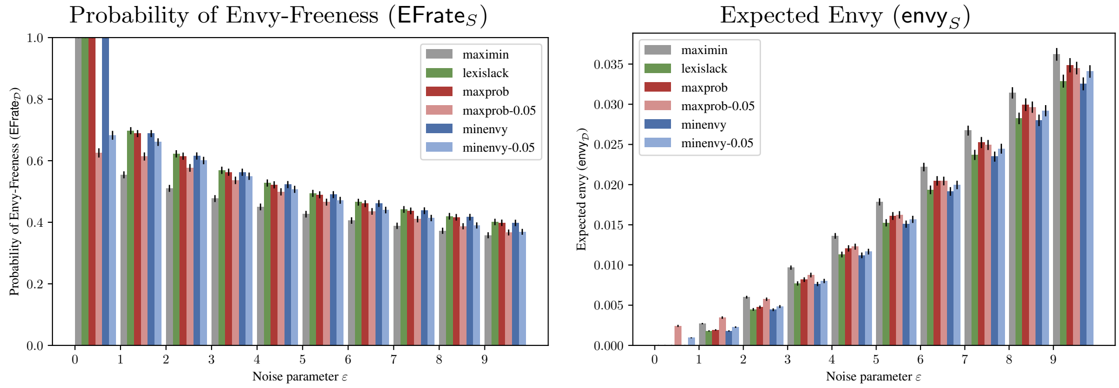


A.3.2. Normal Noise





A.3.3. Biased Normal Noise



A.4. Computation Time

In Figure 3, we show that the computation time of the two probabilistic rules when we vary the noise parameter ε of the underlying distribution. The chart is based on computations involving 1000 instances from spliddit all with $n = 4$, and a sample size of $m = 100$. We can see that with zero noise ($\varepsilon = 0$), computation is extremely fast. This is because all the 100 samples are identical, so the ILP solver can eliminate most constraints as redundant. For positive noise ($\varepsilon \geqslant 0.1$), there is a very slight increase of computation time with increased noise.

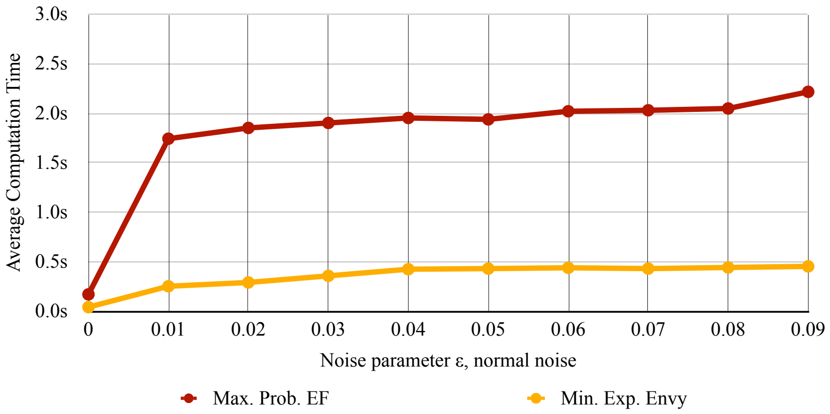


Figure 3: Computation time as a function of the noise parameter

B. ILP Formulations

B.1. Envy-Free Rate

Write $v_{\max} = \max_{i,\ell,r} v_{ir}^{(\ell)}$ (if we use normalized valuations, this value is at most 1). Note that an allocation (σ, p) where $p_r < -v_{\max}$ for some $r \in R$ cannot be envy-free for any of the valuation profiles: If r' is a room with $p_{r'} > 0$, then the person receiving r' values r' at most $v_{\max}$ more than room r, and hence by the price difference will envy the agent receiving room r. So in solving our maximization problem, we can restrict attention to price vectors with $p_r \geqslant -v_{\max}$ for all $r \in R$. Similarly we can assume $p_r \leqslant v_{\max}$.

Using such price vectors, note that the envy between any pair of players is at most $3v_{\max}$. Write $M = 3v_{\max}$.

$$
\begin{aligned}
\max\ & \textstyle\sum_{\ell\in[m]} y_\ell & \\
\text{s.t. } & \textstyle\sum_{r\in R} x_{ir} = 1 & \text{for all } i \in N\\
& \textstyle\sum_{i\in N} x_{ir} = 1 & \text{for all } r \in R\\
& v_{ir}^{(\ell)} - p_r \geqslant v_{ir'}^{(\ell)} - p_{r'} - M(1-y_\ell) - M(1-x_{ir}) & \text{for all } i \in N,\ r, r' \in R,\ \ell \in [m]\\
& \textstyle\sum_{r\in R} p_r = 1 & \\
& -v_{\max} \leqslant p_r \leqslant v_{\max} & \text{for all } r \in R\\
& x_{ir} \in \{0,1\} & \text{for all } i \in N,\ r \in R\\
& y_\ell \in \{0,1\} & \text{for all } \ell \in [m]
\end{aligned}
$$

B.2. Minimize Expected Envy

As in the text, assume that valuations are normalized, and hence restrict attention to reasonable allocations with $-2 \leqslant p_r \leqslant 2$ for all r. Then envy is at most 5. Let $M = 5$.

$$
\begin{aligned}
\min\ & \textstyle\sum_{\ell\in[m]} B_\ell & \\
\text{s.t. } & \textstyle\sum_{r\in R} x_{ir} = 1 & \text{for all } i \in N\\
& \textstyle\sum_{i\in N} x_{ir} = 1 & \text{for all } r \in R\\
& (v_{ir'}^{(\ell)} - p_{r'}) - (v_{ir}^{(\ell)} - p_r) \leqslant B_\ell + M(1-x_{ir}) & \text{for all } i \in N,\ r, r' \in R,\ \ell \in [m]\\
& \textstyle\sum_{r\in R} p_r = 1 & \\
& -2 \leqslant p_r \leqslant 2 & \text{for all } r \in R\\
& x_{ir} \in \{0,1\} & \text{for all } i \in N,\ r \in R\\
& B_\ell \geqslant 0 & \text{for all } \ell \in [m]
\end{aligned}
$$

C. Omitted Proofs

C.1. Proof of Theorem 4.6

Theorem. EF-Rate Maximization with Fixed Assignment *is NP-complete.*

Proof. Membership in NP is clear. We give a reduction from Feedback Arc Set, which can be stated as follows.

Input: Digraph $D = (V, E)$, number B. **Question:** Is there an ordering $(x_1, \dots, x_n)$ of V such that at least B arcs from E point from left to right? (An arc $(x_k \to x_s) \in E$ points from left to right if $k < s$.)

Consider an instance of this problem: Let $D = (V, E)$ be a digraph and let B be a number. We construct a rent division instance where V is both the set of agents and of rooms. Let $\sigma(x) = x$ be the identity room assignment.

Let $\varepsilon = 2/(n(n+1))$, chosen so that $\varepsilon + 2\varepsilon + \cdots + n\varepsilon = 1$.

Label the arc set $E = \{a_1, \ldots, a_m\}$. We define one valuation profile for each arc $a_\ell = (x \to y)$, with

$$v_{xy}^{(\ell)} = 1 + \varepsilon, \qquad v_{zz}^{(\ell)} = 1 \text{ for all } z \in V,$$

and valuation 0 for all unspecified combinations.

We now prove that there is an ordering $(x_1, \ldots, x_n)$ of V with at least B arcs pointing from left to right if and only if there exists a price vector p that makes the identity room assignment σ envy-free in at least B of the valuation profiles.

($\Rightarrow$): Let $(x_1, \ldots, x_n)$ of V be an ordering such that (wlog) the arcs $a_1, \ldots, a_B$ point from left to right.

Consider the price vector $p = (\varepsilon, 2\varepsilon, \ldots, n\varepsilon)$, so room x_i has rent $i \cdot \varepsilon$. This is a valid price vector because it sums up to 1 by choice of ε. We claim that this price vector is envy-free for valuation profiles $v^{(1)}, \ldots, v^{(B)}$. Let $\ell \in [B]$. Since a_ℓ points from left to right, we have $a_\ell = (x_k \to x_s)$ for some $k < s$. First, note that any agent $x_i \neq x_k$ does not envy another agent because x_i values her assigned room x_i at utility 1 higher than other rooms, which avoids envy because room prices differ by less than 1. By the same argument, agent x_k never envies any agent except perhaps x_s. Finally, we check that agent x_k does not envy agent x_s: Note that the rent of room x_s is $s \cdot \varepsilon$, which is at least ε higher than the rent $k \cdot \varepsilon$ of room x_k. Since x_k values room x_s only ε more than her assigned room x_k, she does not envy agent x_s. Thus p is envy-free for valuation profile $v^{(\ell)}$, as required.

($\Leftarrow$): Suppose there is a price vector p that is envy-free for valuation profiles $v^{(1)}, \ldots, v^{(B)}$ (relabeled for convenience). Label the vertices $x_1, \ldots, x_n$ in order of increasing price, i.e., such that $p_{x_1} \leqslant p_{x_2} \leqslant \cdots \leqslant p_{x_n}$ with ties broken arbitrarily. Let $\ell \in [B]$ and consider arc $a_\ell = (x_k \to x_s)$. We show that a_ℓ points from left to right, i.e., $k < s$. As p is envy-free for agent x_k under $v^{(\ell)}$, we have

$$\begin{aligned} v_{x_k,x_s}^{(\ell)} - p_{x_s} \leqslant v_{x_k,x_k}^{(\ell)} - p_{x_k} &\iff 1 + \varepsilon - p_s \leqslant 1 - p_k \\ &\iff p_{x_s} \geqslant p_{x_k} + \varepsilon. \end{aligned}$$

In particular $p_{x_k} < p_{x_s}$. By our choice of ordering, it follows that $k < s$, as required. □

C.2. Proof of Theorem 5.5

Theorem. Expected Envy Minimization *is NP-complete, even for binary valuation profiles.*

Proof. Membership in NP is clear. Reduction from Clique.

Let $G = (V, E)$ be a graph with n vertices and m edges and target clique size k. Set the target envy amount to be $B = m - \binom{k}{2}$. Let M be a large integer, $M > (B+1)^2$. Write $\varepsilon = (B+1)/M$. Our choices of these numbers imply the following estimates which we will need later:

- $M\varepsilon > B$, since $M\varepsilon = B + 1 > B$.
- $\varepsilon(B+1) < 1$, since $\varepsilon(B+1) = (B+1)^2/M < 1$.

The set of agents is V. The set of rooms is $R = \{o_1, \ldots, o_k, d_1, \ldots, d_{n-k}\}$ consisting of k *slot rooms* and $n - k$ dummy rooms. Write $E = \{e_1, \ldots, e_m\}$. We construct a sample S of $m + M$ valuation profiles. For $j \in [m]$, write $e_j = \{u, v\}$; then valuation profile $v^{(j)}$ is defined by

$$v_{i,r}^{(j)} = \begin{cases} 1 & \text{if } i \in \{u, v\} \text{ and } r \in \{o_1, \ldots, o_k\}, \\ 0 & \text{otherwise.} \end{cases}$$

For $j = m+1, \ldots, m+M$, let $v^{(j)}$ be a *uniform profile*:

$$v_{i,r}^{(j)} = 0 \quad \text{for all } i \in V \text{ and } r \in R.$$

($\Rightarrow$): Suppose there is a clique $C \subseteq V$ of size k in G; write $C = \{i_1, \ldots, i_k\}$. We construct an allocation (σ, p) that will be envy-free for B profiles. In the room assignment, we will assign agent i_r to slot room o_r, for $r \in [k]$. The remaining agents can be assigned arbitrarily to dummy rooms. We'll say that each room costs the same rent so $p_r = \frac{1}{n}$ for all $r \in R$.

- For each of the M uniform profiles, $\mathsf{envy}_v(\sigma, p) = 0$.
- For a profile v corresponding to an edge $e_j = \{i_a, i_b\}$ with $i_a, i_b \in C$ (i.e. contained in the clique), $\mathsf{envy}_v = 0$.
- For one of the $m - \binom{k}{2}$ profiles v corresponding to edges not contained in a clique, we have $\mathsf{envy}_v = 1$.

Summing these up, we have $\mathsf{envy}_S(\sigma, p) = m - \binom{k}{2} = B$.

($\Leftarrow$): Suppose there is an allocation (σ, p) with $\mathsf{envy}_S(\sigma, p) \leqslant B$. Let $C \subseteq V$ be the set of k agents/vertices assigned to slot rooms under σ; write $C = \{i_1, \dots, i_k\}$.

First we show that the rents $p = (p_1, \dots, p_n)$ of the rooms are close to uniform, in the sense that $|p_r - p_{r'}| \leqslant \varepsilon$ for all $r, r' \in R$. Assume for a contradiction that there are $r, r' \in R$ with $p_r > p_{r'} + \varepsilon$. Then in each uniform profile, the agent assigned to room r envies the agent assigned to room r' by at least ε, and hence the max envy in a uniform profile is at least ε. Since we have introduced M uniform profiles, it follows that $\mathsf{envy}_S(\sigma, p) \geqslant M\varepsilon > B$, a contradiction.

Now we show that C is a clique. Suppose not. Then there are at least $m - \binom{k}{2} + 1 = B + 1$ edges that are not completely contained in C. For each profile corresponding to such an edge, the agent corresponding to the endpoint not in C envies other agents who are assigned a slot room by at least $1 - \varepsilon$. Hence $\mathsf{envy}_S(\sigma, p) \geqslant (B+1)(1-\varepsilon) = B + 1 - \varepsilon(B+1) > B$, because $\varepsilon(B+1) < 1$. This is a contradiction. $\square$