
Online certification of preference-based fairness for personalized recommender systems

Virginie Do^{1,2} Sam Corbett-Davies² Jamal Atif¹ Nicolas Usunier²

Abstract

We propose to assess the fairness of personalized recommender systems in the sense of *envy-freeness*: every (group of) user(s) should prefer their recommendations to the recommendations of other (groups of) users. Auditing for envy-freeness requires probing user preferences to detect potential blind spots, which may deteriorate recommendation performance. To control the cost of exploration, we propose an auditing algorithm based on pure exploration and conservative constraints in multi-armed bandits. We study, both theoretically and empirically, the trade-offs achieved by this algorithm.

1. Introduction

Recommender systems shape the information available to us and the opportunities we have access to. They help us prioritize content from news outlets and social networks, sort job offers, or find new people to connect with – from professional contacts to potential life partners. Substantial work has been done to audit online ranking or recommender systems (Sweeney, 2013; Hannak et al., 2014; Lambrecht & Tucker, 2019). The goal has mostly been to assess parity criteria, for instance, whether housing recommendations are independent of the declared gender of the user. These audits focus only on the distribution of recommended items without regard to individual users’ preferences. Their application to personalized recommender systems is thus limited to sanity checks such as the one described above, i.e., assessing whether recommendations are sensitive to an (inferred) protected group membership.

In this paper, we study algorithms that audit recommender systems from the perspective of preference-based fairness (Zafar et al., 2017; Ustun et al., 2019; Kim et al., 2019). We focus on recommendations that are *envy-free*, which means

that users prefer their recommendations to the recommendations of other users. We also propose an extension of envy-free recommendations to groups of users rather than individuals, which is non-trivial because there is no single “group preference” nor “group recommendation”.

Envy-freeness was originally studied in the domain of fair allocation (Foley, 1967). For recommender systems, it considers personalization fair as long as it is faithful to the underlying preferences of users. Moreover, unlike approaches based on parity measures, or the requirement that groups enjoy similar recommendation performance, envy-freeness is always compatible with optimal personalized recommendations. Yet, envy-freeness is not a performance criterion since giving the same recommendation to everyone is envy-free. In practice, recommender systems do personalize their predictions, but they are necessarily suboptimal because of the scarcity of per-user data. In that context, envy-freeness leaves enough freedom to accommodate every users’ preferences, while precluding the unfair situation where users looking for the same opportunities are given recommendations of different quality.

The main challenge of auditing for envy-freeness is that it requires access to user preferences, which are only partially observed since users only interact with recommended items. Thus, to identify the recommender system’s potential blind spots, the auditor needs an active exploration process that recommends items which would not have been recommended otherwise. To make such an exploration possible, we consider a scenario where the auditor is allowed to replace a user’s recommendations with the recommendations that another user would have received in the same context. Envy, or the absence thereof, is estimated by suitably choosing whose recommendations should be shown to whom.

We first provide a formal analysis of envy-free recommender systems, including a non-trivial definition for groups. Then, we cast the problem of auditing for (group) envy-freeness as a new variant of pure exploration problems in bandits (Audibert & Bubeck, 2010; Bubeck et al., 2009). To alleviate the potential negative side effect of exploration, namely that users may be given suboptimal recommendations, we follow the literature on conservative exploration (Wu et al., 2016; Garcelon et al., 2020) and propose an algorithm that

¹LAMSADE, PSL, Université Paris Dauphine, France

²Facebook AI. Correspondence to: Virginie Do <virginie.do@dauphine.eu>.

provably maintains, throughout the course of the audit, a performance close to the audited system. We study, formally and experimentally, the trade-offs between the three main criteria of the audit: the desired statistical confidence, the duration of the audit, and the per-user cost of exploration.

We discuss the related work in Sec. 2. Envy-free and group envy-free recommender systems are studied in Sec. 3 and 4. The auditing algorithm and its theoretical guarantees are presented in Sec. 5, and the experimental analysis in Sec. 6.

2. Related work

Fair recommendation The domain of fairness in machine learning is organized along two orthogonal axes. The first axis is whether fairness is oriented towards groups defined by protected attributes (Barocas & Selbst, 2016), or rather oriented towards individuals (Dwork et al., 2012). The second axis is whether fairness is a question of *parity* (predictions [or prediction errors] should be the same between groups or individuals) (Corbett-Davies & Goel, 2018), or *preference-based* (predictions are allowed to be different as long as they faithfully reflect the preferences of all parties) (Zafar et al., 2017; Kim et al., 2019; Ustun et al., 2019). Our work starts from the perspective of envy-freeness, which follows the preference-based approach and is oriented towards individuals. We also consider an extension to groups, though these need not be defined by protected attributes.

The literature on fair recommender systems covers two problems: *auditing* existing systems, and *designing* fair recommendation algorithms. Most of the *auditing* literature focused on group parity in recommendations (Hannak et al., 2014; Lambrecht & Tucker, 2019), and equal user satisfaction (Mehrotra et al., 2017; Ekstrand et al., 2018), while our audit for envy-freeness focuses on whether personalized results are aligned with (unknown) user preferences. On the *designing* side, Patro et al. (2020); Ilvento et al. (2020) cast fair recommendation as an allocation problem, with criteria akin to envy-freeness. They do not address the partial observability of preferences, so they cannot guarantee user-side fairness without an additional certificate that the estimated preferences effectively represent the true user preferences. Our work is thus complementary to theirs.

While we emphasized fairness for users, recommender systems are multi-sided (Burke, 2017; Patro et al., 2020), thus fairness can also be oriented towards recommended items (Celis et al., 2017; Biega et al., 2018; Geyik et al., 2019).

Multi-arm bandits In *pure exploration* bandits (Bubeck et al., 2009), an agent has to identify a specific set of arms after exploring as quickly as possible, without performance constraints. Our setting is close to threshold bandits (Locatelli et al., 2016) where the goal is to find arms with better

performance than a given baseline. Outside pure exploration, in the *regret minimization* setting, conservative exploration (Wu et al., 2016) enforces the anytime average performance to be not too far worse than that of a baseline arm.

In our work, the baseline is *unknown* – it is the current recommender system – and the other “arms” are other users’ policies. The goal is to make the decision as to whether an arm is better than the baseline, while not deteriorating performance compared to the baseline. We thus combine pure exploration and conservative constraints.

Existing work on fairness in exploration/exploitation (Joseph et al., 2016; Jabbari et al., 2017; Liu et al., 2017) is different from ours because unrelated to personalization.

Envy in fair division Envy-freeness was first studied in social choice (specifically, in fair resource allocation) (Foley, 1967). Our setting is different because: a) the same item can be given to an unrestricted number of users, and b) true user preferences are unknown. There are several notions of group envy-freeness in fair division, either considering every subset of individuals (e.g., Berliant et al. (1992)), or pre-defined groups (e.g., Manurangsi & Suksompong (2017)), but they are conceptually different from ours.

3. Envy-free recommendations

3.1. Framework

There are M users, and we identify the set of users with $[M] = \{1, \dots, M\}$. A personalized recommender system has one stochastic recommendation policy π^m per user m . We denote by $\pi^m(a|x)$ the probability of recommending item $a \in \mathcal{A}$ for user $m \in [M]$ in context $x \in \mathcal{X}$. We assume that \mathcal{X} and \mathcal{A} are finite to simplify notation, but this has no impact on the results. We consider a synchronous setting where at each time step, the recommender system observes a context $x_t^m \sim q^m$ for each user, selects an item $a_t^m \sim \pi^m(\cdot|x_t^m)$ and observes reward $r_t^m \sim \nu^m(a_t^m|x_t^m) \in [0, 1]$. We denote by $\rho^m(a|x)$ the expected reward for user m and item a in context x , and, for any recommendation policy π , $u^m(\pi)$ is the utility of m for π :

$$\begin{aligned} u^m(\pi) &= \mathbb{E}_{x \sim q^m} \mathbb{E}_{a \sim \pi(\cdot|x)} \mathbb{E}_{r \sim \nu^m(a|x)} [r] \\ &= \sum_{x \in \mathcal{X}} \sum_{a \in \mathcal{A}} q^m(x) \pi(a|x) \rho^m(a|x) \end{aligned} \quad (1)$$

We assume that the environment is *stationary*: the context and reward distributions q^m and ν^m , as well as the policies π^m are fixed. Even though in practice policies evolve as they learn from user interactions and user needs change over time, we leave the study of non-stationarities for future work. The stationary assumption approximately holds when these changes are slow compared to the time horizon of the audit, which is reasonable when significant changes in user needs

or recommendation policies take e.g., weeks. Our approach applies when items a are single products as well as when items are structured objects such as rankings. Examples of (context x , item a) pairs include: x is a query to a search engine and a is a document or a ranking of documents, or x is a song chosen by the user and a a song to play next or an entire playlist. Remember, our goal is *not* to learn the user policies π^m , but rather to audit existing π^m s for fairness.

3.2. ϵ -envy-free recommendations

The preference-based approach to fair machine learning stipulates that it is fair to apply different policies to different individuals or groups as long as it is in everyone’s interest. Following this principle, we consider that the personalization of recommendations is fair only if it better accommodates individuals’ preferences. The fairness criterion associated to preference-based fairness is *envy-freeness* (Foley, 1967; Balcan et al., 2018; Ustun et al., 2019; Kim et al., 2019). Adapting the previous definitions to the context of personalized recommendation, we consider that a system is fair as long as each user (weakly) prefers their recommendations to the recommendations of other users:

Definition 3.1. Let $\epsilon \geq 0$. A recommender system is ϵ -envy-free if: $\forall m, n \in [M] : u^m(\pi^n) \leq \epsilon + u^m(\pi^m)$.

Envy-freeness neither guarantees the parity of recommendations, nor equal user satisfaction. In particular, since a non-personalized recommender system¹ is always envy-free, a system that only follows the preferences of one (group of) users can be envy-free, even if it is unfair according to other standards. As usual in the assessment of fairness, several diagnostics should be applied for a full assessment. Parity and equal satisfaction are straightforward to estimate since they do not require any form of exploration, which is why we focus in this paper only on how to audit for envy-freeness.

Among these criteria, envy-freeness is the only one compatible with optimal recommendations (see below). For instance, equal user satisfaction is impossible to achieve when some users have intrinsically more noisy preferences. Violating parity or equality of satisfaction is only an indication that the system *may* fail, while violating envy-freeness is an indication that the system *does* fail. It is thus a much stronger signal than the other diagnostics.

Envy-freeness vs. optimality certificates Let $\pi^{m,*} \in \arg\max_{\pi} u^m(\pi)$ denote an optimal recommendation policy for m . Then the *optimal recommender system* $(\pi^{m,*})_{m \in M}$ is *envy-free* since: $u^m(\pi^{m,*}) = \max_{\pi} u^m(\pi) \geq u^m(\pi^{n,*})$.

To understand the differences between a certificate of envy-freeness and a certificate of optimality, let us denote by $\Pi^* = \{\pi : \exists u \text{ satisfying (1), } \pi \in \arg\max_{\pi'} u(\pi')\}$ the set

¹i.e., $\forall m, n, \pi^m = \pi^n$

of potentially optimal policies. If the set of users policies approximately covers the set of potentially optimal policies Π^* , then an envy-free system is also optimal.² In practice, the space of optimal policies is much larger than the number of users (for instance, there are $|\mathcal{A}|^{|\mathcal{X}|}$ optimal policies in our setting), so that auditing for envy is tractable in cases where auditing for optimality is not.

Sources of envy Real-world recommender systems cannot learn optimal recommendation policies due to the lack of per-user data. They must rely on strong modeling assumptions with methods such as low-rank matrix factorization (Koren et al., 2009). This leads to envy when the system misrepresents the preferences of some users while matching those of users with similar tastes. Another source of envy is when the measurement of relevant side information varies across groups (Suresh & Guttag, 2019). Many hybrid recommender systems indeed rely on user-side data such as listening history on a music streaming platform. Systematic envy arises if this information is unevenly collected across users, for example if the platform’s inventory does not cover niche music tastes. We show experimentally in Sec. 6.1 and App. B.2 how *mispecified models* (e.g. with a rank that is too low) and *measurement bias* can create envy.

3.3. Compatibility with item-side fairness

Envy-freeness is a user-centric notion. Towards multisided fairness (Burke, 2017), we analyze the compatibility of envy-freeness with item-side fairness criteria for rankings from Singh & Joachims (2018), based on sensitive categories of items (denoted $\mathcal{A}_1, \dots, \mathcal{A}_S$). *Parity of exposure* prescribes that for each user, the exposure of an item category should be proportional to the number of items in that category. In *Equity of exposure*³, the exposure of item categories should be proportional to their average relevance to the user.

The optimal policies under parity and equity of exposure constraints, denoted respectively by $(\pi^{m,\text{par}})_{m=1}^M$ and $(\pi^{m,\text{eq}})_{m=1}^M$, are defined given user m and context x as:

$$\begin{aligned} (\text{parity}) \quad \pi^{m,\text{par}}(\cdot|x) &= \arg\max_{\substack{p: \mathcal{A} \rightarrow [0,1] \\ \sum_a p(a)=1}} \sum_{a \in \mathcal{A}} p(a) \rho^m(a|x) \\ \text{u.c. } \forall s \in [S], \sum_{a \in \mathcal{A}_s} p(a) &= \frac{|\mathcal{A}_s|}{|\mathcal{A}|}. \quad (2) \end{aligned}$$

Optimal policies under equity of exposure are defined simi-

²Formally, let $D(\pi, \pi')$ such that $|u(\pi) - u(\pi')| \leq D(\pi, \pi')$. It is easy to see that if $\max_{\pi \in \Pi^*} \min_{m \in M} D(\pi, \pi^m) \leq \tilde{\epsilon}$, then ϵ -envy-freeness implies $\epsilon + \tilde{\epsilon}$ -optimality.

³Singh & Joachims (2018) use a different terminology of demographic parity (resp. disparate treatment) constraints for what we call parity (resp. equity) of exposure. Our use of “equity” in that context follows Biega et al. (2018).

larly⁴, but the constraints are $\forall s, \sum_{a \in \mathcal{A}_s} p(a) = \frac{\sum_{a \in \mathcal{A}_s} \rho^m(a|x)}{\sum_{a \in \mathcal{A}} \rho^m(a|x)}$.

We show their relation to envy-freeness:

Proposition 1. *With the above notation:*

- the policies $(\pi^{m,\text{par}})_{m=1}^M$ are envy-free, while
- the policies $(\pi^{m,\text{eq}})_{m=1}^M$ are not envy-free in general.

Optimal recommendations under parity of exposure are envy-free because the parity constraint (2) is the same for all users. Given two users m and n , $\pi^{m,\text{par}}$ is optimal for m under (2) and $\pi^{n,\text{par}}$ satisfies the same constraint, so we have $u^m(\pi^{m,\text{par}}) \geq u^m(\pi^{n,\text{par}})$.

In contrast, the optimal recommendations under equity of exposure are, in general, not envy-free. A first reason is that less relevant item categories reduce the exposure of more relevant categories: a user who prefers item a but who also likes item b from another category envies a user who only liked item a . Note that *amortized* versions of the criterion and other variants considering constraint averages over user/contexts (Biega et al., 2018; Patro et al., 2020) have similar pitfalls unless envy-freeness is explicitly enforced, as in Patro et al. (2020) who developed an envy-free algorithm assuming the true preferences are known. For completeness, we describe in App. A a second reason why equity of exposure constraints create envy.

3.4. Auditing scenario

The auditing scenario is the following: at each time step t , the auditor chooses to either (a) give the user a “normal” recommendation, or (b) explore user preferences by giving the user a recommendation from another user. This scenario has the advantage of lightweight infrastructure requirements, since the auditor only needs to query another user’s policy, rather than implementing a full recommender system within the operational constraints of the platform. Moreover, this interface is sufficient to estimate envy because envy is defined based on the performance of other user’s policies. This type of internal audit (Raji et al., 2020) requires more access than usual external audits that focus on recommendation parity, but this is necessary to explore user preferences.

We note the auditor must make sure that this approach follows the relevant ethical standard for randomized experiments in the context of the audited system. The auditor must also check that using other user’s recommendation policies does not pose privacy problems. From now on, we assume these issues have been resolved.

⁴The original criterion (Singh & Joachims, 2018, Eq. 4) would be written in our case as $\forall s, s' \in [S], \frac{1}{|\mathcal{A}_s|} \sum_{a \in \mathcal{A}_s} p(a) = \frac{1}{|\mathcal{A}_{s'}|} \sum_{a \in \mathcal{A}_{s'}} p(a)$, which is equivalent to (2). A similar remark holds for the equity constraint.

4. Group envy-free recommendations

In this section, we present an extension of envy-free recommendation to groups rather than individuals. The motivation for this extension is twofold. First, it generalizes the existing definitions of preference-based fairness (which are oriented towards protected groups rather than individuals (Zafar et al., 2017; Ustun et al., 2019)) to personalized predictions. Second, as we shall see in the next section, assessing group-level envy requires less exploration per user as groups become larger. Thus, in practice, it is likely that only group-level certificates have manageable sample complexity.

From now on, we assume we are given a partition of the users g_1, \dots, g_G , i.e., $\bigcup_{i \in [G]} g_i = [M]$ and $g_i \cap g_j = \emptyset$ when $i \neq j$. We define the group utility μ^i for g_i as the average utility of users in g_i :

$$\mu^i = \frac{1}{|g_i|} \sum_{m \in g_i} u^m(\pi^m). \quad (3)$$

There are two desiderata for a definition of group envy-free recommendations:

- following our scenario, we can only query other users’ policies to perform exploration actions.
- group envy-freeness must relax envy-freeness, in the sense that an ϵ -envy-free recommender system must be ϵ -group envy-free for every partition of users.

The notion of envy-free recommendations described in the previous section uses $u^m(\pi^n)$, the utility of user m for user n ’s policy. The challenge is to define an analogous quantity between groups, which we solve using what we call *matching policies*. Given two groups i and k , we compute new policies π_k^m for every user m in g_i by “reallocating” them the recommendations made to the users of g_k . To solve point *i*) above, this reallocation is done by considering matching policies that are mixtures of users’ policies: For every groups g_i, g_k and every user $m \in g_i$:

$$\exists (\gamma_{m,n})_{n \in g_k} \text{ s.t. } \pi_k^m = \sum_{n \in g_k} \gamma_{m,n} \pi^n \quad (4)$$

$$\text{where } \forall n \in g_k, \gamma_{m,n} \geq 0 \text{ and } \sum_{n \in g_k} \gamma_{m,n} = 1$$

Notice that our recommendation policies are stochastic, so $\gamma_{m,n}$ reweights the probabilities to recommend each item. We study several possible definitions for $\gamma_{m,n}$ later in this section. The advantage of these matching policies is that the auditor can obtain a recommendation from π_k^m by first sampling n from $(\gamma_{m,n})_{n \in g_k}$ and then querying π^n , which fits our auditing scenario (point *i*) above).

With this constraint in mind, we can define ϵ -group envy-freeness using matching policies:

Definition 4.1. Given $i, k \in [G]$ and matching policies $(\pi_k^m)_{m \in g_i}$, the utility of g_k with respect to g_i is

$$\mu_k^i = \frac{1}{|g_i|} \sum_{m \in g_i} u^m(\pi_k^m). \quad (5)$$

Given $\epsilon \geq 0$, the recommender system is ϵ -group-envy-free if

$$\forall i, k \in [G] : \mu_k^i \leq \mu_i^i + \epsilon. \quad (6)$$

Group envy-freeness is equivalent to envy-freeness when each group is a singleton. We use μ_k^i instead of μ^i in (6) to make it comparable to μ_k^i . The difference between the two is that μ_k^i uses the matching policies π_k^m instead of the usual recommendation policies π^m , and we may have $\pi_k^m \neq \pi^i$ depending on the matching rule. The important result is that point *ii*) is satisfied when $\mu_k^i = \mu^i$:

Proposition 2. If the mixture weights in (4) are such that $\forall i \in [G], \mu_k^i = \mu^i$, then an ϵ -envy-free recommender system is ϵ -group envy-free for every partition $(g_i)_{i \in [G]}$.

Proof. Since $(\pi^m)_{m=1}^M$ are ϵ -envy-free, for every $i, k \in [G]$ and users $m \in g_i, n \in g_k$, we have, $u^m(\pi^n) \leq u^m(\pi^m) + \epsilon$. Since u^m is linear in the policy, we have

$$u^m(\pi_k^m) = \sum_{n \in g_k} \gamma_{m,n} u^m(\pi^n) \leq \sum_{n \in g_k} \gamma_{m,n} (u^m(\pi^m) + \epsilon).$$

Since $\sum_{n \in g_k} \gamma_{m,n} = 1$, we have $u^m(\pi_k^m) \leq u^m(\pi^m) + \epsilon$. Summing over $m \in g_i$, we obtain $\mu_k^i \leq \mu^i + \epsilon = \mu_i^i + \epsilon$. \square

We now give two examples of matching policies to illustrate the importance of Proposition 2.

Average group policy A first possible definition for a matching policy is to take the average policy over a group. Let $i, k \in [G]$ and $m \in g_i$. The average group policy of g_k for user m is $\pi_k^m = \frac{1}{|g_k|} \sum_{n \in g_k} \pi^n$, i.e., $\gamma_{m,n} = \frac{1}{|g_k|}$ for all $n \in g_k$ in (4). Unfortunately, this simple matching policy does not satisfy desideratum *ii*):

Proposition 3. With average group policies, there exist envy-free recommender systems that are not group envy-free.

The reason for Prop. 3 is that when a group contains users with different policies, the average policy represents no one's preferences and $\mu_k^i \ll \mu^i$, violating the condition of Prop. 2. In some cases, this leads to detecting group envy even though there is no individual envy. A minimal example of such cases is presented in Table 1.

Mixtures from optimal transport In order to guarantee $\mu_k^i = \mu^i$, we propose to define the mixture weights $\gamma_{m,n}$ for $m \in g_i$ and $n \in g_k$ using optimal transport (Villani, 2008). The distance used to match two users m, n is the supremum

		item 1	item 2	μ^i	μ_k^1	μ_k^2
g_1	ρ^1, π^1 ρ^2, π^2	1, 1 0.5, 0	0, 0 1, 1	1		
g_2	ρ^3, π^3 ρ^4, π^4	1, 1 1, 1	0, 0 0, 0	1		
$(k=1)$	π_1^m	0.5	0.5		0.625	0.5
$(k=2)$	π_2^m	1	0		0.75	1

Table 1. Example where average group policies fail on desideratum *ii*). The policies π^m are optimal for utilities ρ^m , so the recommendations are envy-free. However, because users in g_1 have different optimal policies, the average policy is bad for both ($\mu_1^1 \ll \mu^1$). Using average group policies, group envy is undesirably detected.

(over contexts) of the variational between policies $c_{m,n} = \sup_{x \in \mathcal{X}} \|\pi^m(\cdot|x) - \pi^n(\cdot|x)\|_1$. Without prior knowledge, policies are the only available user information to the auditor. The matching could also use a task-specific distance (Dwork et al., 2012), if additional user information from the platform or guidelines from a policymaker are accessible.

Let \mathcal{P} be the transportation polytope between the uniform distributions over g_i and g_k , i.e., each $\gamma \in \mathcal{P}$ is an $|g_i| \times |g_k|$ matrix with non-negative entries, rows summing to $\frac{1}{|g_i|}$ and columns summing to $\frac{1}{|g_k|}$. The transported policies are then defined as:

Definition 4.2. Let $i, k \in [G]$. Define $\tilde{\gamma}^{i,k}$ and γ as:

$$\tilde{\gamma}^{i,k} = \operatorname{argmin}_{\tilde{\gamma} \in \mathcal{P}} \sum_{\substack{m \in g_i \\ n \in g_k}} \tilde{\gamma}_{m,n} c_{m,n} \quad \forall m \in g_i, \gamma_{m,n} = |g_i| \tilde{\gamma}_{m,n}^{i,k}.$$

For $m \in g_i$, the transported policy is $\pi_k^m = \sum_{n \in g_k} \gamma_{m,n} \pi^n$.

Notice that γ defines a proper mixture since $\sum_{n \in g_k} \gamma_{m,n} = 1$.

Proposition 4. With transported policies, an ϵ -envy-free recommender system is ϵ -group envy-free for every $(g_i)_{i \in [G]}$.

Prop. 4 then follows from Prop. 2, since with our definition of c , matched users have the same policies, so that $\mu^i = \mu_k^i$.

In App. C, we describe ways to construct groups in practice.

5. Certificates of envy-freeness

We now turn to our auditing algorithm for envy-freeness. We directly focus on groups and only consider transported policies. For every group g_i , the auditor must estimate whether all other groups k satisfy $\mu_k^i \leq \mu^i + \epsilon$. To estimate μ_k^i , we obtain samples by making recommendations using the transported policy π_k^m and observing the reward, for every $m \in g_i$. The remaining challenge is to choose which group k to sample at each time step while not deteriorating the user experience too much.

Algorithm 1: OCEF algorithm. ξ_t (line 4) evaluates the conservative exploration constraint and is defined in (8). Values for $\beta_k(t)$ and confidence bounds $\underline{\mu}_k$ and $\bar{\mu}_k$ are given in Lemma 3.

input : Confidence parameter δ , conservative exploration parameter α , envy parameter ϵ
output : ENVY or ϵ -NO ENVY

```

1  $S_0 \leftarrow [K]$  // all arms except 0
2 for  $t=1, \dots$  do
3   Choose  $\ell_t$  from  $S_{t-1}$  // e.g., unif. sample
4   if  $\beta_0(t-1) > \min_{k \in S_{t-1}} \beta_k(t-1)$  or  $\xi_t < 0$  then  $k_t \leftarrow 0$ 
5   else  $k_t \leftarrow \ell_t$ 
6    $r_t \leftarrow \text{PULL-ARM}(k_t)$ 
   // update conf. intervals with Lem. 3
7    $S_t \leftarrow \{k \in S_{t-1} : \bar{\mu}_k(t) > \underline{\mu}_0(t) + \epsilon\}$ 
8   if  $\exists k \in S_t, \underline{\mu}_k(t) > \bar{\mu}_0(t)$  then return ENVY
9   if  $S_t = \emptyset$  then return  $\epsilon$ -NO ENVY
10 end
    
```

We first describe the auditing algorithm when estimating envy for a single target group g_i . For a full audit of a recommender system, we apply the algorithm to all groups in parallel. To simplify notation, we drop all superscripts i : we use g and μ_k instead of g_i and μ_k^i . We also re-index groups from 0 to $K = G - 1$, where index 0 represents the target group (μ_0 is thus the utility of the group for the audited recommender system). Because the audit is a special form of bandit problem, following the bandit literature, an index of a group is called an *arm*, and arm 0 is the *baseline*.

Objectives and evaluation metrics We present our algorithm OCEF (Online Certification of Envy-Freeness) in the next subsection. Given $\epsilon > 0$ and $\alpha \geq 0$, OCEF returns either ENVY or ϵ -NO-ENVY and has two objectives:

1. Correctness: if OCEF returns ENVY, then $\exists k, \mu_k > \mu_0$.
 If OCEF returns ϵ -NO-ENVY then $\max_{k \in [K]} \mu_k \leq \mu_0 + \epsilon$.
2. Recommendation performance: during the audit, OCEF must maintain a fraction $1-\alpha$ of the baseline performance. Denoting by $k_s \in \{0, \dots, K\}$ the arm (group index) chosen at round s , this requirement is formalized as a conservative exploration constraint (Wu et al., 2016):

$$\forall t, \frac{1}{t} \sum_{s=1}^t \mu_{k_s} \geq (1-\alpha)\mu_0. \quad (7)$$

We focus on the *fixed confidence* setting, where given a confidence parameter $\delta \in (0, 1)$ the algorithm provably satisfies both objectives with probability $1 - \delta$. In addition, there are two criteria to assess an online auditing algorithm:

1. Duration of the audit: the number of time-steps before the algorithm stops.
2. Cost of the audit: the cumulative loss of rewards incurred. Denoting the duration by τ , the cost is $\tau\mu_0 - \sum_{s=1}^{\tau} \mu_{k_s}$.

It is possible that the cost is negative when there is envy. In that case, the audit increased recommendation performance by finding better recommendations for the group.

We note the asymmetry in the return statements of the algorithm: ENVY does not depend on ϵ . This asymmetry is necessary to obtain finite worst-case bounds on the duration and the cost of audit, as we see in Theorem 1. We now describe the algorithm, and then provide theoretical guarantees for the objectives and evaluation measures.

5.1. The OCEF algorithm

OCEF is described in Alg. 1. It maintains confidence intervals on arm performances $(\mu_k)_{k=0}^K$. Given the confidence parameter δ , the lower and upper bounds on μ_k at time step t , denoted by $\underline{\mu}_k(t)$ and $\bar{\mu}_k(t)$, are chosen so that with probability at least $1 - \delta$, we have $\forall k, t, \mu_k \in [\underline{\mu}_k(t), \bar{\mu}_k(t)]$. In the algorithm, $\beta_k(t) = (\bar{\mu}_k(t) - \underline{\mu}_k(t))/2$. As Jamieson et al. (2014), we use anytime bounds inspired by the law of the iterated logarithm. These are given in Lem. 3 in App. G.

OCEF maintains an active set S_t of all arms in $[K]$ (i.e., excluding the baseline) whose performance are not confidently less than $\mu_0 + \epsilon$. It is initialized to $S_0 = [K]$ (line 1). At each round t , the algorithm selects an arm $\ell_t \in S_t$ (line 3). Then, depending on the state of the conservative exploration constraint (described later), the algorithm pulls k_t , which is either ℓ_t or the baseline (lines 4-6). After observing the reward r_t , the confidence interval of μ_{ℓ_t} is updated, and all active arms that are confidently worse than the baseline plus ϵ are de-activated (line 7). The algorithm returns ENVY if an arm k is confidently better than the baseline (line 8), returns ϵ -NO-ENVY if there are no more active arms, (line 9) or continues if neither of these conditions are met.

Conservative exploration To deal with the conservative exploration constraint (7), we follow (Garcelon et al., 2020). Denoting $A_t = \{s \leq t : k_s \neq 0\}$ the time steps at which the baseline was not pulled, we maintain a confidence interval such that with probability $\geq 1 - \delta$, we have $\forall t > 0, |\sum_{s \in A_t} (\mu_{k_s} - r_s)| \leq \Phi(t)$. The formula for Φ is given in Lem. 4 in App. G. This confidence interval is used to estimate whether the conservative constraint (7) is met at round t as follows. First, let us denote by $N_k(t)$ the number of times arm k has been pulled until t , and notice that (7) is equivalent to $\sum_{s \in A_t} \mu_{k_s} - ((1-\alpha)t - N_0(t))\mu_0 \geq 0$. After choosing ℓ_t (line 3), we use the lower bound on $\sum_{s \in A_t} \mu_{k_s}$ and the upper bound for μ_0 to obtain a conservative estimate

of (7). Using $\tau = t - 1$, this leads to:

$$\xi_t = \sum_{s \in A_\tau} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(\tau) + (N_0(\tau) - (1 - \alpha)t)\bar{\mu}_0(\tau). \quad (8)$$

Then, as long as the confidence intervals hold, pulling ℓ_t does not break the constraint (7) if $\xi_t \geq 0$. The algorithm thus pulls the baseline arm when $\xi_t < 0$. To simplify the theoretical analysis, OCEF also pulls the baseline if it does not have the tightest confidence interval (lines 4-6).

PULL-ARM OCEF is an exploration algorithm which only chooses a group index. Once a group has been chosen, PULL-ARM (line 6 of Alg. 1) chooses the effective recommendations to make to users. The PULL-ARM routine is described in Alg. 2 (App. F). If OCEF chooses the baseline arm, then every user m in the audited group receives a recommendation from π^m as usual. If OCEF chooses $k \neq 0$, then for every user m in the audited group, the recommendation uses the transported policy π_k^m defined in 4.2. It first samples $n \sim (\gamma_{m,n'})_{n' \in g_k}$, and then queries n 's policy π^n .

5.2. Analysis

The main theoretical result of the paper is the following:

Theorem 1. *Let $\epsilon \in (0, 1]$, $\alpha \in (0, 1]$, $\delta \in (0, \frac{1}{2})$ and $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$ and $h_k = \max(1, (|g|\eta_k)^{-1})$. Using $\underline{\mu}$, $\bar{\mu}$ and Φ given in Lemmas 3 and 4 (App. G), OCEF achieves the following guarantees with probability $\geq 1 - \delta$:*

- OCEF is correct and satisfies the conservative constraint on the recommendation performance (7).
- The duration is in $O\left(\sum_{k=1}^K \frac{h_k \log\left(\frac{K \log(K h_k / \delta \eta_k)}{\delta}\right)}{\min(\alpha \mu_0, \eta_k)}\right)$.
- The cost is in $O\left(\sum_{k: \mu_k < \mu_0} \frac{(\mu_0 - \mu_k) h_k}{\eta_k} \log\left(\frac{K \log(K h_k / \delta \eta_k)}{\delta}\right)\right)$.

The important problem-dependent quantity η_k is the gap between the baseline and other arms k . It is asymmetric depending on whether the arm is better than the baseline ($\mu_k - \mu_0$) or the converse ($\mu_0 - \mu_k + \epsilon$) because the stopping condition for ENVY does not depend on ϵ . This leads to a worst case that only depends on ϵ , since $\eta_k = \max(\mu_k - \mu_0, \mu_0 - \mu_k + \epsilon) \geq \frac{\epsilon}{2}$, while if the condition was symmetric, we would have possibly unbounded duration when $\mu_k = \mu_0 + \epsilon$ for some $k \neq 0$. Overall, ignoring log terms, we conclude that when $\alpha \mu_0$ is large, the duration is of order $\sum_k \frac{1}{|g|\eta_k^2}$ and the cost is of order $\sum_k \frac{1}{|g|\eta_k}$. This becomes $\sum_k \frac{1}{\alpha \mu_0 |g|\eta_k}$ and $\sum_k \frac{1}{|g|\eta_k}$ when $\alpha \mu_0$ is small compared to η_k . This means that the conservative constraint has an impact mostly when it is relatively strict. It also means that

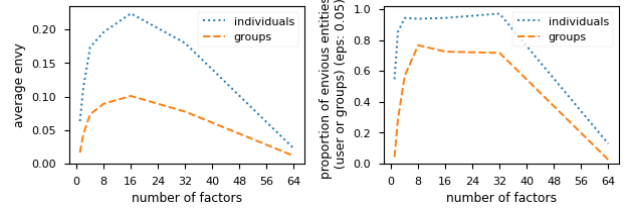


Figure 1. Effect of model misspecification on MovieLens, with 121 groups defined by intersecting age and occupation.

when either $\alpha \mu_0 \ll \eta_k$ or $\eta_k^2 \ll \eta_k$ the cost can be small even when the duration is fairly high. In all cases, it is critical to note that both the cost and duration, which are per-user quantities, decrease linearly with the group size $|g|$. Thus assessing group envy-freeness becomes more sample efficient (per user) as groups become larger.

Full audit To audit for the full system, we apply OCEF to all groups simultaneously. By the union bound, using $\delta' = \frac{\delta}{G}$ instead of δ in OCEF's confidence intervals, the guarantees of Theorem 1 hold simultaneously for all groups.

6. Experiments

We present two experiments describing sources of envy (Sec. 6.1) and one experiment evaluating the auditing algorithm OCEF on a movie recommendation task (Sec. 6.2).

We consider two measures of the degree of envy for groups (with equivalent notions for individuals). Denoting $\Delta^i = \max(\max_{k \in [G]} \mu_k^i - \mu_i^i, 0)$, these are:

- the average envy experienced by groups: $\frac{1}{G} \sum_{i \in [G]} \Delta^i$,
- the proportion of ϵ -envious groups: $\frac{1}{G} \sum_{i \in [G]} \mathbb{1}_{\{\Delta^i > \epsilon\}}$.

6.1. Envy from model misspecification on MovieLens

We demonstrate that envy arises from a standard recommendation model when the modeling assumptions are too strong. We simulate a movie recommendation task using the MovieLens-1M dataset (Harper & Konstan, 2015), which contains ratings of movies by real users together with demographic information. We extract a 2000×2000 user \times items matrix, keeping users and items with the most ratings, and fill in missing entries using a popular matrix completion algorithm⁵. This completed matrix serves as ground truth user preferences. Recommendations use low-rank matrix completion (Bell & Sejnowski, 1995) on a training sample of 20% of the ratings, where the rated items are sampled

⁵Using <http://github.com/gbolmier/funk-svd>.

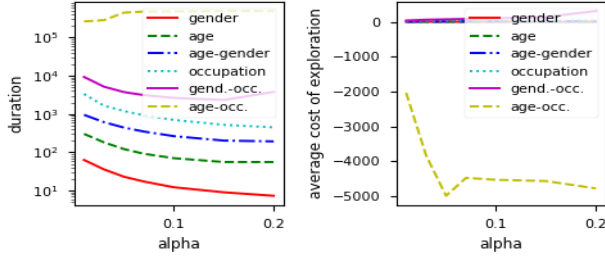


Figure 2. Scaling w.r.t. α on MovieLens experiments. There are 2 gender, 7 age, and 21 occupation groups.

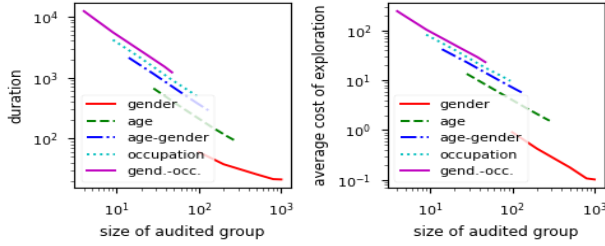


Figure 3. Scaling w.r.t. group size on MovieLens experiments. Larger groups allow for shorter durations.

uniformly at random. We generate binary ratings using a Bernoulli distribution with expectation given by our ground truth. Recommendations use a *softmax* policy over the predicted ratings. We study other policies in App. B.1.

We define user groups by MovieLens demographic attributes and their intersection. In particular, there are 2 groups for gender, 7 for age, 21 for occupation, 14 for age-gender, 42 for gender-occupation, and 120 for age-occupation.

In Figure 1, we consider individuals and age-occupation groups. We observe that with one latent factor there is no envy. This is because all users receive the same recommendations since matrix completion is then equivalent to a popularity-based recommender system. With enough latent factors, preferences are properly captured by the model and the degree of envy tends to zero. For intermediate number of latent factors however, envy is visible.

6.2. Evaluation of the audit algorithm on MovieLens

Our goal is now to answer the following for OCEF: (1) What is the interplay between the required number of samples per user, the cost of exploration and the conservative exploration parameter? (2) In practice, how does the audit duration change as the number of groups or groups' sizes changes?

The task is to certify the (absence of) envy of the *argmax* policy for different types of groups on MovieLens. We use 50 latent factors. There is envy for 9 age-occupation groups, and no envy for the other group configurations.

Fig. 2 shows the result of OCEF on the various types of groups, where the curves are averaged on all groups, except for age-occupation groups where the average is over envious groups only (and averaged over 20 trials, with $\delta = \epsilon = 0.05$). The curves show that for gender-occupation and age-occupation, duration is minimal for a non-trivial α . This is because when α is large, all arms are pulled as much as the baseline, so their confidence intervals are similar. When α decreases, the baseline is pulled more, which reduces the length of the relevant confidence intervals $\beta_0(t) + \beta_k(t)$ for all arms k . This, in turn, shortens the audit because non-baseline arms are more rapidly discarded or declared better. When α becomes too small, however, the additional pulls of the baseline have no effect on $\beta_0(t) + \beta_k(t)$ because it is dominated by $\beta_k(t)$, so the duration only increases. This subtle phenomenon is not captured by our analysis (Th. 1), because the ratios $\beta_0(t)/\beta_k(t)$ are difficult to track formally.

The sign of the cost of exploration depends on whether there is envy. For envious age-occupation groups, exploration is beneficial to the users and so the cost is negative. On all other partitions however, the cost is positive. The cost of exploration increases with α and is closest to 0 when α becomes small because then $\beta_0(t) + \beta_k(t)$ is the smallest possible for a given number of pulls of k .

The actual orders of magnitude of the audit duration in Fig. 2 seem large with more granular groups (e.g., $10^3 - 10^4$ time steps for gender-occupation), but this is because the groups are small since there are only 2000 users in total. However, as seen in Th. 1, larger groups allow for shorter audits and smaller cost of exploration per user. To give a more complete picture with larger user databases, we artificially generate larger groups by randomly duplicating users. Fig. 3 presents the duration of the experiments as a function of the average size of the group. We see that, as expected from the theory, durations decrease linearly with the absolute group size, while also increasing with the number of groups. Looking at age-gender, we see that group sizes of 150 users already allow for audits of 150 time steps, which is promising for online services with large user databases. We give more details in App. D.2 and present qualitatively similar results on another dataset in App. E.

7. Conclusion

We proposed the audit of recommender systems for user-side fairness with the criterion of envy-freeness, and discussed a generalized definition of group envy-freeness. The auditing problem requires an explicit exploration of user preferences, which leads to a formulation as a bandit problem with conservative constraints. We presented an algorithm for this problem and analyzed its performance experimentally.

References

- Audibert, J.-Y. and Bubeck, S. Best arm identification in multi-armed bandits. 2010.
- Balcan, M.-F., Dick, T., Noothigattu, R., and Procaccia, A. D. Envy-free classificatoion. *arXiv preprint arXiv:1809.08700*, 2018.
- Barocas, S. and Selbst, A. D. Big data’s disparate impact. *Calif. L. Rev.*, 104:671–769, 2016.
- Bell, A. J. and Sejnowski, T. J. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995.
- Berliant, M., Thomson, W., and Dunz, K. On the fair division of a heterogeneous commodity. *Journal of Mathematical Economics*, 21(3):201–216, 1992.
- Biega, A. J., Gummadi, K. P., and Weikum, G. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, pp. 405–414, 2018.
- Bubeck, S., Munos, R., and Stoltz, G. Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pp. 23–37. Springer, 2009.
- Burke, R. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002.
- Burke, R. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017.
- Celis, L. E., Straszak, D., and Vishnoi, N. K. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840*, 2017.
- Chevalleyre, Y., Endriss, U., and Maudet, N. Distributed fair allocation of indivisible goods. *Artificial Intelligence*, 242:1–22, 2017.
- Corbett-Davies, S. and Goel, S. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*, 2018.
- Dwork, C., Hardt, M., Pitassi, T., Reingold, O., and Zemel, R. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pp. 214–226. ACM, 2012.
- Ekstrand, M. D., Tian, M., Azpiazu, I. M., Ekstrand, J. D., Anuyah, O., McNeill, D., and Pera, M. S. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on Fairness, Accountability and Transparency*, pp. 172–186. PMLR, 2018.
- Flamary, R. and Courty, N. Pot python optimal transport library. *GitHub: <https://github.com/rflamary/POT>*, 2017.
- Foley, D. K. Resource allocation and the public sector. 1967.
- Freedman, D. A. On tail probabilities for martingales. *the Annals of Probability*, pp. 100–118, 1975.
- Garcelon, E., Ghavamzadeh, M., Lazaric, A., and Pirotta, M. Conservative exploration in reinforcement learning. In *International Conference on Artificial Intelligence and Statistics*, pp. 1431–1441, 2020.
- Geyik, S. C., Ambler, S., and Kenthapadi, K. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2221–2231, 2019.
- Hannak, A., Soeller, G., Lazer, D., Mislove, A., and Wilson, C. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*, pp. 305–318, 2014.
- Harper, F. M. and Konstan, J. A. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015.
- He, R., Kang, W.-C., and McAuley, J. Translation-based recommendation. In *Proceedings of the eleventh ACM conference on recommender systems*, pp. 161–169, 2017.
- Hoeffding, W. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.
- Ilvento, C., Jagadeesan, M., and Chawla, S. Multi-category fairness in sponsored search auctions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 348–358, 2020.
- Jabbari, S., Joseph, M., Kearns, M., Morgenstern, J., and Roth, A. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 1617–1626. JMLR. org, 2017.
- Jamieson, K. and Nowak, R. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pp. 1–6. IEEE, 2014.

- Jamieson, K., Malloy, M., Nowak, R., and Bubeck, S. *lil'ucb*: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pp. 423–439, 2014.
- Järvelin, K. and Kekäläinen, J. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002.
- Joseph, M., Kearns, M., Morgenstern, J. H., and Roth, A. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, pp. 325–333, 2016.
- Kearns, M., Neel, S., Roth, A., and Wu, Z. S. Preventing fairness gerrymandering: Auditing and learning for subgroup fairness. In *International Conference on Machine Learning*, pp. 2564–2572, 2018.
- Kim, M. P., Korolova, A., Rothblum, G. N., and Yona, G. Preference-informed fairness. *arXiv preprint arXiv:1904.01793*, 2019.
- Koren, Y., Bell, R., and Volinsky, C. Matrix factorization techniques for recommender systems. *Computer*, 42(8): 30–37, 2009.
- Lambrech, A. and Tucker, C. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, 65(7): 2966–2981, 2019.
- Liu, Y., Radanovic, G., Dimitrakakis, C., Mandal, D., and Parkes, D. C. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*, 2017.
- Locatelli, A., Gutzeit, M., and Carpentier, A. An optimal algorithm for the thresholding bandit problem. *arXiv preprint arXiv:1605.08671*, 2016.
- Manurangsi, P. and Suksompong, W. Asymptotic existence of fair divisions for groups. *Mathematical Social Sciences*, 89:100–108, 2017.
- Mehrotra, R., Anderson, A., Diaz, F., Sharma, A., Wallach, H., and Yilmaz, E. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, pp. 626–633, 2017.
- Patro, G. K., Biswas, A., Ganguly, N., Gummadi, K. P., and Chakraborty, A. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In *Proceedings of The Web Conference 2020*, pp. 1194–1204, 2020.
- Raji, I. D., Smart, A., White, R. N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., and Barnes, P. Closing the ai accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 33–44, 2020.
- Singh, A. and Joachims, T. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2219–2228. ACM, 2018.
- Suresh, H. and Gutttag, J. V. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, 2019.
- Sweeney, L. Discrimination in online ad delivery. *Queue*, 11(3):10, 2013.
- Ustun, B., Liu, Y., and Parkes, D. Fairness without harm: Decoupled classifiers with preference guarantees. In *International Conference on Machine Learning*, pp. 6373–6382, 2019.
- Villani, C. *Optimal transport: old and new*, volume 338. Springer Science & Business Media, 2008.
- Wu, Y., Shariff, R., Lattimore, T., and Szepesvári, C. Conservative bandits. In *International Conference on Machine Learning*, pp. 1254–1262, 2016.
- Zafar, M. B., Valera, I., Rodriguez, M., Gummadi, K., and Weller, A. From parity to preference-based notions of fairness in classification. In *Advances in Neural Information Processing Systems*, pp. 228–238, 2017.

		item cat. 1		item cat. 2		utilities	
(item idx)		1	2	3	4	u^1	u^2
(rewards)	ρ^1	1	0	0.8	0.7		
	ρ^2	0.8	0.7	1	0		
(policies)	$\pi^{1,\text{eq}}$	0.4	0	0.6	0	0.88	0.92
	$\pi^{2,\text{eq}}$	0.6	0	0.4	0	0.92	0.88

Table 2. Example where the optimal recommendations under item-side equity of exposure constraints are not user-side fair because both users envy each other. There are 4 items, 2 item categories and 2 users. User 1 envies user 2 since $u^1(\pi^{2,\text{eq}}) > u^1(\pi^{1,\text{eq}})$. Also, $u^2(\pi^{1,\text{eq}}) > u^2(\pi^{2,\text{eq}})$.

A. More on equity of exposure

We remind the definition of optimal policies with equity of exposure constraints from Section 3.3:

$$\begin{aligned}
 (\text{equity}) \quad \pi^{m,\text{eq}}(\cdot|x) &= \underset{\substack{p: \mathcal{A} \rightarrow [0,1] \\ \sum_a p(a)=1}}{\operatorname{argmax}} \sum_{a \in \mathcal{A}} p(a) \rho^m(a|x) \\
 \text{u.c. } \forall s \in [S], \sum_{a \in \mathcal{A}_s} p(a) &= \frac{\sum_{a \in \mathcal{A}_s} \rho^m(a|x)}{\sum_{a \in \mathcal{A}} \rho^m(a|x)}
 \end{aligned}$$

The constraints should be ignored when $\sum_{a \in \mathcal{A}} \rho^m(a|x) = 0$.

Following Proposition 1 from Section 3.3, we describe here a second source of envy when using optimal policies with equity of exposure constraints. By the linearity of the optimization problem for $\pi^{m,\text{eq}}$, the policy assigns to the best item in a category the exposure of the entire category. It implies that categories with high average relevance have more exposure than categories with few but highly relevant items. Table 2 gives an example with two users and two categories of items where both users envy each other with the optimal recommendations under equity of exposure constraints.

In some degenerate cases though, equity of exposure policies are envy-free.

Lemma 2. *If for all contexts $x \in \mathcal{X}$, each user $m \in [M]$ only likes a single item category \mathcal{A}_{s_m} , i.e. $\forall a \in \mathcal{A} \setminus \mathcal{A}_{s_m}, \rho^m(a|x) = 0$, then the policies $(\pi^{m,\text{eq}})_{m=1}^M$ are envy-free.*

Proof. We set contexts x aside to simplify notation, but the generalization is straightforward.

We actually prove a stronger result than the lemma: if each user m only likes a single item, then $(\pi^{m,\text{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$, where $\pi^{m,*}$ is the optimal unconstrained policy for m .

Let $a_s^m = \operatorname{argmax}_{a \in \mathcal{A}_s} \rho^m(a)$ be the favorite item in category \mathcal{A}_s for user m , then the optimal equity of exposure

constrained policies has the following analytical expression:

$$\forall s \in S, \forall a \in \mathcal{A}_s, \quad \pi^{m,\text{eq}}(a) = \mathbb{1}_{\{a=a_s^m\}} \frac{\sum_{a' \in \mathcal{A}_s} \rho^m(a')}{\sum_{a' \in \mathcal{A}} \rho^m(a')},$$

and we thus have:

$$u^m(\pi^{m,\text{eq}}) = \sum_{s \in [S]} \rho^m(a_s^m) \frac{\sum_{a \in \mathcal{A}_s} \rho^m(a)}{\sum_{a \in \mathcal{A}} \rho^m(a)}. \quad (9)$$

If each user $m \in [M]$ only likes a single item category $s_m \in [S]$, i.e. $\forall a \in \mathcal{A} \setminus \mathcal{A}_{s_m}, \rho^m(a) = 0$, then $\frac{\sum_{a \in \mathcal{A}_{s_m}} \rho^m(a)}{\sum_{a \in \mathcal{A}} \rho^m(a)} = \mathbb{1}_{\{s=s_m\}}$.

Then $u^m(\pi^{m,\text{eq}}) = \rho^m(a_{s_m}^m) = \max_{a \in \mathcal{A}} \rho^m(a)$.

Then $\pi^{m,\text{eq}}$ is the optimal unconstrained policy for user m , meaning the whole system is envy-free (cf. Sec 3.2).

From Eq. 9, we actually note that $(\pi^{m,\text{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$ if and only if each user m equally values their favorite items in each category they like, i.e. $\forall m, \exists \kappa > 0, \forall s \in S, \rho^m(a_s^m) > 0 \Rightarrow \rho^m(a_s^m) = \kappa$.

□

B. Sources of envy

In this section, we first list a few possible sources of envy in recommender systems. Then we describe experiments⁶ which showcase two of these sources, namely model misspecification (App. B.1) and measurement bias (App. B.2).

Examples of sources of envy

Model misspecification Recommender systems often rely on strong modeling assumptions and multi-task learning, with methods such as low-rank matrix factorization (Koren et al., 2009). The limited capacity of the models (e.g., a rank that is too low) or incorrect assumptions might leave aside users with less common preference patterns. Appendix B.1 gives a more detailed example than Sec. 6 on a simulated movie recommendation task.

Misaligned incentives A recommender system might have incentives to recommend some items to specific users, e.g., sponsored content. Envy appears when there is a mismatch between users who like these items and users to whom they are recommended.

Measurement bias Many hybrid recommender systems rely on user interactions together with user-side data (Burke,

⁶For all our experiments, we used Python and a machine with Intel Xeon Gold 6230 CPUs, 2.10 GHz, 1.3 MiB of cache.

2002). This includes side-information such as browsing history on third-party, partner websites. Envy arises in these settings if there is measurement bias (Suresh & Gutttag, 2019), e.g., if the side information is unevenly collected for all users. Group envy arises when the information is missing in correlation to relevant social groups (e.g., browsing patterns are different per groups and partners are aligned with the patterns of a few groups only). Appendix B.2 provides an example environment where group envy arises.

Operational constraints Regardless of incentives, recommendations might need to obey additional constraints. As described in Proposition 1, the item-side fairness constraint of equity of exposure is an example of possible source of (user-side) envy.

In the following, we present examples of environments where envy holds at the individual or group level, based on two of the previous sources of envy. The experiments for model misspecification (App. B.1) extend our simulations from Section 6.1 to various types of policies, and demonstrates how envy is observed at the level of individuals and small groups, on a typical movie recommendation task. For measurement bias (App. B.2), we describe a synthetic controlled experiment where envy systematically arises for groups of any sizes.

In these experiments, we measure envy for groups based on the quantity:

$$\Delta^i = \max(\max_{k \in [G]} \mu_k^i - \mu_i^i, 0).$$

In line with (Chevalleyre et al., 2017), we consider two ways of measuring the degree of envy, stated in the language of groups (with equivalent notions for individuals):

- the average envy experienced by groups: $\frac{1}{G} \sum_{i \in [G]} \Delta^i$,
- the proportion of ϵ -envious groups: $\frac{1}{G} \sum_{i \in [G]} \mathbb{1}_{\{\Delta^i > \epsilon\}}$.

B.1. Experiments with envy from model misspecification

We simulate a movie recommendation task using the MovieLens-1M dataset (Harper & Konstan, 2015), which contains 1 million ratings from approximately 6000 users and 4000 movies. We use demographic attributes to group users by their gender (2 groups), age (7 groups), occupation (21 groups), intersection of gender-age (14 groups) and gender-occupation (42 groups), age-occupation (121 groups). We use movie genres to create item categories, keeping only the first genre if several are reported (18 categories).

We extract a 2000×2000 user \times items matrix, keeping users and items with the most ratings, and fill in missing

entries using a popular matrix completion algorithm⁷. This completed matrix serves as ground truth user preferences. We then simulate a recommender system’s estimation of preferences using low-rank matrix completion (Bell & Sejnowski, 1995) on a training sample of 20% of the ratings, where the rated items are sampled uniformly at random for each user. We generate binary ratings using a Bernoulli distribution with expectation given by our ground truth. This procedure is the same as the one used in our experiments presented in Section 6.1. More details on hyperparameters and methods can be found in Appendix D.2.

The first type of recommendation policies we use is the unconstrained optimal policy (see Section 3.2) computed on estimated preferences. It is called the *argmax* policy because it recommends the single estimated preferred item of a user. The second type of policy, the *PAR* policy, is the optimal policy under parity of exposure constraints (see Section 3.3) computed on estimated preferences as well. In that case, we use movie genres to create item categories, keeping only the first genre if several are reported (18 categories).

For both policies above, the recommendation task is to recommend a single item. We also add a ranking task. The *ranking* policy returns the ordered list of top-10 items according to estimated preferences. We use a standard ranking metric to evaluate the utility of a ranking, the Discounted Cumulative Gain (Järvelin & Kekäläinen, 2002). Given a user m and a policy π , it is computed by:

$$\text{DCG}(\pi) = \sum_{j=1}^{10} \frac{\rho^m(a_j^\pi)}{\log(1+j)}$$

where a_j^π is the j -th item in the ranking given by π , and $\rho^m(a_j^\pi)$ is the ground truth value of that item for user m .

In Figures 4, 5, 6, the variation of envy with the number of latent factors is qualitatively similar to the experiment with softmax policy shown in Sec. 6.1, with a peak of the degree of envy for a nontrivial number of latent factors.

In the next experiment, we describe a controlled experiment where envy arises systematically for a disadvantaged group, regardless of its size.

B.2. Experiments with group envy from measurement bias

We consider an instance of measurement bias (Suresh & Gutttag, 2019), where the observation of relevant information varies across groups. Our example is a streaming platform whose inventory barely covers the music tastes of some specific group. The system misses a large part of their true listening history, which results in systematically noisier recommendations of new items for the disadvantaged group.

⁷Using <http://github.com/gbolmier/funk-svd>.

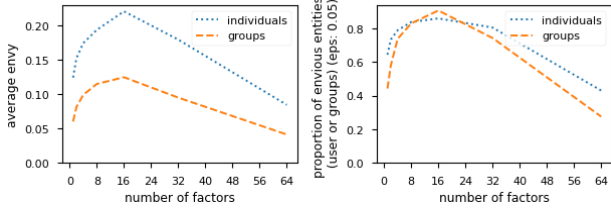


Figure 4. Model mispecification on MovieLens with Armax policy. Groups are defined by age-occupation.

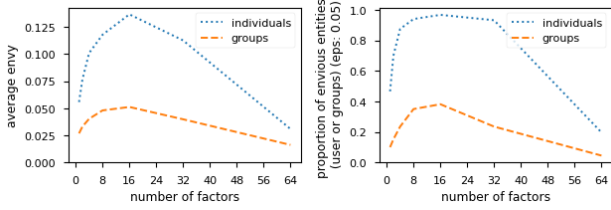


Figure 5. Model mispecification on MovieLens with Ranking policy. Groups are defined by age-occupation.

Formal model Formally, we consider the item set \mathcal{A} (e.g., new music tracks to recommend) and a set of documents \mathcal{D} (e.g., tracks in the inventory) of size D that users visited in the past. There are no contexts in this model.

We first generate ground-truth user preferences $p^m(a)$ by taking the dot product between a user embedding \mathbf{v}^m and an item embedding \mathbf{e}_a in a latent space of dimension d . These embeddings are drawn uniformly at random from unit vectors with positive coordinates of dimension d . Ratings sampled from this ground truth follow a Bernoulli distribution with parameter $\sigma(\alpha \langle \mathbf{v}^m, \mathbf{e}_a \rangle - t)$ where t is a threshold and σ is the sigmoid function. Therefore, the true preferences over new items in \mathcal{A} follow the same distribution across groups.

We correlate the true listening history of m in group g_i with its group membership as follows. Let $\mathcal{H}_1 = \{1, \dots, \lfloor \frac{d}{2} \rfloor\}$ and $\mathcal{H}_2 = \{\lfloor \frac{d}{2} \rfloor + 1, \dots, d\}$, and let τ_1, τ_2 be group-specific parameters, with $\tau_1 > \tau_2$. For $h \in \mathcal{D}$, let \mathbf{e}_h be a document embedding generated similarly to the item embeddings, and let

$$X_h^m \sim \text{Pois}(\tau_i \langle \mathbf{v}^m, \mathbf{e}_h \rangle) \mathbf{1}_{\{h \in \mathcal{H}_i\}}.$$

This means that the observations $X^m \in \mathbb{R}^D$ of the listening history for each user m are groupwise orthogonal, and that the amount of available listening counts varies by group (as controlled by τ_i).

Based on these biased observations, the recommender system learns a simple rating model by linear regression in the document space, and the system computes a stochastic policy that is a *softmax* distribution over these estimated ratings,

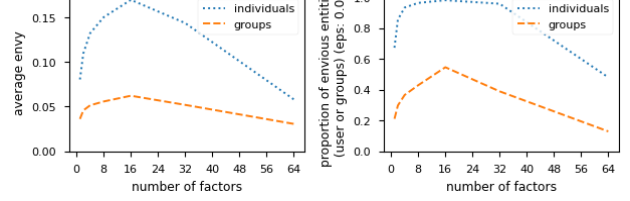


Figure 6. Model mispecification on MovieLens with Parity of exposure. Groups are defined by age-occupation.

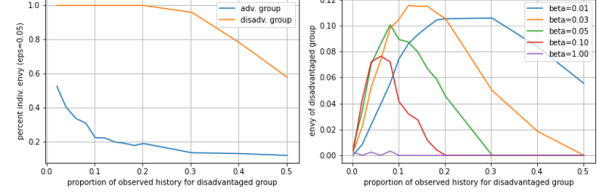


Figure 7. Envy caused by measurement bias, as a function of the proportion of observed data for the disadvantaged groups.

with parameter β to control the softmax temperature:

$$\hat{r}^m(a) = \frac{X^m}{\|X^m\|_2} \hat{\mathbf{E}}_a^T, \quad \pi^{m, \text{softmax}}(a) = \sigma(\beta \hat{r}^m(\cdot))_a$$

Due to the lack of historical data for the disadvantaged group g_2 whose representation is orthogonal to that of g_1 , the system is likely to incorrectly predict the best item for any user m in g_2 , whereas their counterpart n with similar preferences in the advantaged group g_1 is accurately given their favorite item. Therefore, any m in g_2 is likely to envy their alter ego n in g_1 .

Simulation We consider 2 groups of 1000 users each (i.e., $M = 2000$), $|\mathcal{A}| = 200$ new items to recommend and an inventory of $|\mathcal{D}| = 200$ documents. For the ground truth ratings, we use a latent space of dimension $d = 10$ and set the parameters $t = \frac{1}{d}$, $\alpha = 30d$. For the correlation of the history with groups, we set parameter $\tau_1 = 50$ and we vary τ_2 in $\{0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50\}$. For the softmax policy, we vary the thermodynamic β in $\{0.01, 0.025, 0.05, 0.1, 1\}$. Ground truth ratings for the items \mathcal{A} are sampled uniformly random, such that the linear regression uses a 20%/20%/60% train/valid/test split. The softmax policy $\pi^{m, \text{softmax}}$ is restricted to items in the test subset.

Recall that true user preferences over new items follow the same distribution across groups, whereas groups have orthogonal listening histories. Therefore, group g_2 is disadvantaged since less data is available for them. Fig. 7 plots individual and group envy as a function of the proportion of observed items for the disadvantaged group. We observe

that the proportion of ϵ -envious individual users in the disadvantaged group g_2 decreases when a larger proportion of the listening history is available.

We now analyze group envy for the disadvantaged group. We observe that when no history is observed for the disadvantaged group, there is no envy. This is because in the absence of data, the predictions are random, and so users are matched randomly from g_2 to g_1 . When some history is observed but in small proportion, group envy is highest, and it decreases as a larger proportion of the history is revealed. Moreover, the softmax parameter β has an effect on group envy: when β increases, the softmax policy is closer to an argmax policy. Therefore, the ground cost used when computing transported policies becomes uninformative: users are matched uniformly at random between g_2 and g_1 , hence our inability to observe envy when $\beta = 1$.

C. Groups in practice

Here we describe a favorable property of homogeneous groups and its implication for constructing groups in practice.

When policies in groups are homogeneous, then the average group policy and the transported policy are similar, so ϵ -envy-freeness translates to ϵ' -group envy-freeness for average group policies as well, with $\epsilon' \approx \epsilon$. More interestingly, when we have prior knowledge that users in all groups have homogeneous preferences and homogeneous policies, the reciprocal of Proposition 4 from Sec. 4 holds for both average and transported policies:

Proposition 5. *Let $\epsilon, \tilde{\epsilon} > 0$, and assume that for all groups $i \in [G]$ and all pairs of users m, n in the same group g_i , we have*

$$\begin{aligned} \sup_{x \in \mathcal{X}} \|\pi^m(\cdot|x) - \pi^n(\cdot|x)\|_1 &\leq \tilde{\epsilon} \text{ and} \\ \sup_{x \in \mathcal{X}} \|\rho^m(\cdot|x) - \rho^n(\cdot|x)\|_1 &\leq \tilde{\epsilon}. \end{aligned}$$

Then, with either average group policies or transported group policies, ϵ -group envy-freeness implies $(\epsilon + 4\tilde{\epsilon})$ -envy-freeness.

The result is natural since when all groups have users with homogeneous preferences and policies, groups and users are a similar entity as far as the assessment of envy-freeness is concerned. The proof is straightforward and omitted. The interest of this remark lies in a better understanding of how to construct groups in practice, since homogeneity in policies and preferences impact how much a certificate of ϵ -group envy-freeness should be interpreted as a certificate of ϵ' -envy-freeness with $\epsilon' \approx \epsilon$.

In the favorable case where we have strong prior knowledge on user preferences, building groups by clustering by user

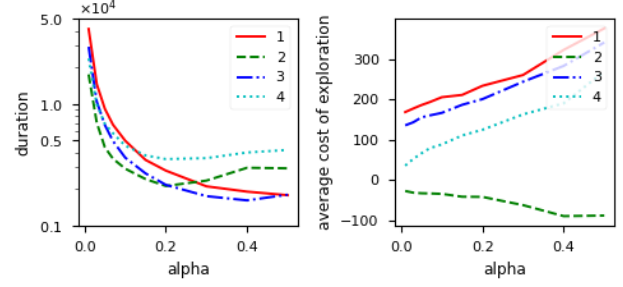


Figure 8. Effect of the conservative exploration parameter α on the duration and cost of auditing on Bandit experiments.

preferences allows the conditions of Proposition 5 to hold (assuming the system meets the minimal requirement of similar policies given similar preferences). The number of users in each cluster controls the trade-off between the strength of the guarantee (in terms of $\epsilon + 4\tilde{\epsilon}$ of Prop. 5⁸) and the number of required exploration steps per user. In that case, the certification of envy-freeness aims at certifying that preferences are properly represented in the recommendations.

In many cases, however, the auditor does not have prior knowledge on user preferences. An audit for group envy-freeness in this case will likely only be able to detect envy when some groups receive systematically worse recommendations than other groups. In the literature on fairness in machine learning, groups are most often defined by identity of sensitive attributes such as gender, ethnicity, age (Barocas & Selbst, 2016), as well as intersections of them (Kearns et al., 2018), because these are dimensions of frequent systematic undesirable bias. Thus, groups that only differ by a value of these attributes should be compared.

D. OCEF experiments

D.1. Bandit experiments

We performed experiments on toy bandit environments to assess the performance of our algorithm OCEF on various configurations, which were also considered in (Jamieson & Nowak, 2014). The four bandits instances have 10 arms. They are Bernoulli variables with means equal to

- 1) $\mu_0 = 0.6$ and $\mu_k = 0.3$ for $k \in [9]$,
- 2) $\mu_0 = 0.3$, $\mu_1 = 0.6$ and $\mu_k = 0.3$ for $k = 2..9$,
- 3) $\mu_k = 0.7 - 0.7 * (\frac{k}{10})^{0.6}$, $k = 0, \dots, 9$, and the baseline is μ_0 ,
- 4) same as 3), but permuting μ_0 and μ_1 .

⁸ $\tilde{\epsilon}$ is typically larger as soon as groups contain more users.

Fig. 8 shows the result of applying OCEF on the various configurations, where we set $\delta = \epsilon = 0.05$ and report results averaged over 100 trials. We observe clear tendencies similar to those of the MovieLens experiments presented in Section 6.2, although the exact sweet spots in terms of α depends on the specific configuration.

The cost of exploration follows similar patterns as in the MovieLens experiments. In Prob. 2, the baseline has the worst performance, so exploration is beneficial to the user and the cost is negative. On the other hand, for instance in Prob. 4, the cost is close to 0 when α is very small and increases with α . It is the case where the baseline is not the best arm but is close to it, and there are many bad arms. When the algorithm is very conservative, bad arms are discarded rapidly thanks to the good estimation of the baseline performance. In this “low-cost” regime however, the audit is significantly longer.

We show additional results when varying δ in Figure 9. Results are averaged over 100 simulations and the conservative exploration parameter is set to $\alpha = 0.05$. The duration decreases as δ increases, i.e. a lower confidence certificate requires fewer samples per user. The duration for Problem 1 is longer than for the other instances. This is because with α set to 0.05 and the baseline mean being much higher than non-baseline arms, the conservative constraint 7 enforces many pulls of the baseline, since each exploration round is very costly. As a consequence, too little data is collected on the non-baseline arms to conclude that they are below $\mu_0 + \epsilon$. Since all non-baseline arms have equal means, the size of the active set remains the same for a long time, while in Problem 3, where the baseline is also the best arm, arms are eliminated one at a time.

We show how OCEF scales with the number of arms in Figure 10, for fixed values $\alpha = \delta = \epsilon = 0.05$. We set $K_{\max} = 100$ and define 4 instances as in the list above, except that $K = K_{\max}$ instead of $K = 9$. We run OCEF on the instances $\mu_{0:K'}$ and vary the value of $K' \leq K_{\max}$. The duration increases for all problems, and the slope depends on the gaps between μ_0 and the μ_k .

D.2. MovieLens experiments

We sample 20% of the entries uniformly at random for each user to be observed, and the observed value is drawn from a Bernoulli distribution with mean given by the ground truth value. Matrix factorization is run with 50 latent factors, using a Python implementation⁹ of a popular matrix completion algorithm. Hyperparameter search is conducted for the regularization parameter chosen in [0.01, 0.1, 1, 10]. Notice that we aim at estimating envy of the resulting recommender system, so the exact training

⁹Using <https://github.com/gbolmier/funk-svd>

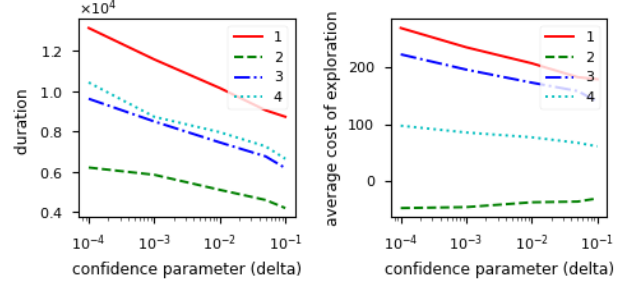


Figure 9. Effect of the confidence parameter δ on the duration and cost on 4 different bandit instances.

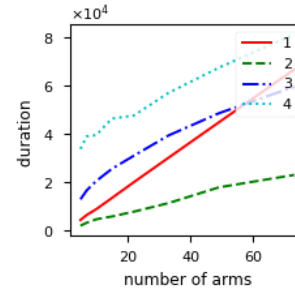


Figure 10. Effect of the number of arms on the duration on 4 different bandit instances.

method is not important as long as it is reasonable. We use an argmax policy that recommends the best item according to the estimated preferences. The transported policies are computed using Python Optimal Transport Library (Flamary & Courty, 2017).

For $\alpha = \epsilon = 0.05$, the effect of varying the confidence parameter δ is shown in Figure 11. Consistently with our upper bound (Theorem 1), the number of samples collected per user (left) decreases as δ increases. The sample complexity required for the partition gender-occupation (42 groups) is orders larger than for the partition by gender (2 groups). This highlights again the trade-off between the granularity of the partition of users in group and the duration of the audit. The cost per user (right) follows the same decreasing trend.

E. Additional experiments

We present a batch of experiments with different settings than the main one presented in Section 6. Since all results are qualitatively similar, we did not include them in the core paper. The goal is to confirm the following claim: the sparsity of the training set only affects the quality of the recommender system, and not the auditing process.

Using the same protocol, we repeated the experiments

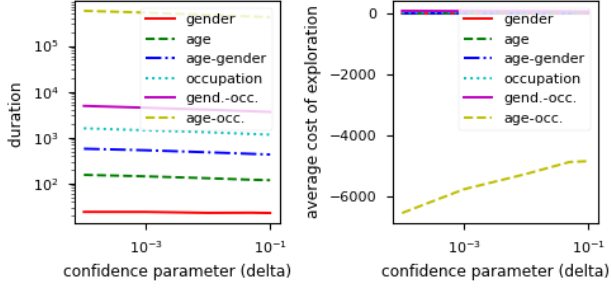


Figure 11. Scaling w.r.t. δ on MovieLens experiments.

from Section 6 using 1) subsampled training sets from MovieLens-1M with different levels of sparsity and 2) an additional ratings dataset, Google Local ratings¹⁰.

Precisely, to study the effect of sparsity, we subsampled training sets with 10% and 5% of the MovieLens “ground truth” ratings (obtained with matrix completion). As in our main experiments, these training sets are given as input to the FunkSVD algorithm, and the recommender system uses a softmax policy over predicted ratings. We remind that in Sec. 6, the same protocole was used with a training set made of 20% of the ratings.

Then, to compare our results on MovieLens with another dataset, we use the Google Local ratings dataset from (He et al., 2017). It includes reviews and ratings from 11,4M users about 3,1M businesses. As in (Patro et al., 2020), we restrict the dataset to users and businesses located in New York and with at least 10 reviews. We draw a random sample of 2,000 users among the remaining ones. With the same protocole as for MovieLens, we generate “ground truth” ratings for each user-item pair with standard matrix completion. This results in a full rating matrix with 2,000 users and 855 items. As before, we simulate a recommender system by training Funk-SVD on a subsample of 20% of the ratings, and using a softmax policy over predicted ratings. In the absence of demographic attributes, we focus on individual envy-freeness.

For the experiments on envy from misspecification (as in Sec. 6.1), Figures 12, 13, 14 show that the degree of envy varies similarly w.r.t. the number of latent factors across experimental settings. The only difference is the amplitude of envy.

Then, we evaluate the auditing algorithm (as in Sec. 6.2) on each of these environments, and observe that the performance of OCEF is not affected. Fig. 15, 16 show that the results are qualitatively similar when recommendations are generated from sparser rating matrices extracted from

¹⁰Available at <https://cseweb.ucsd.edu/~jmcauley/datasets.html>

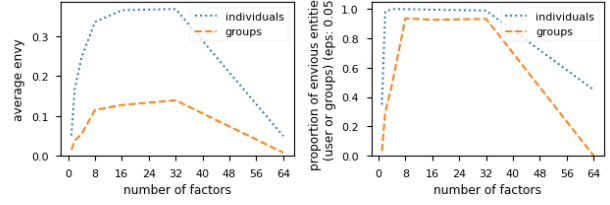


Figure 12. Model misspecification on MovieLens with Softmax policy, trained on 5% of the ratings. Groups are defined by age-occupation.

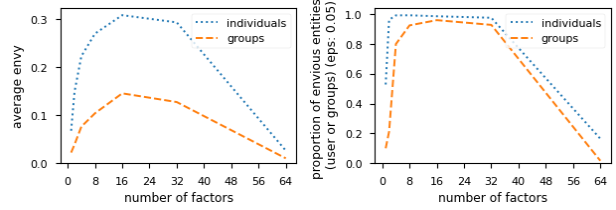


Figure 13. Model misspecification on MovieLens with Softmax policy, trained on 10% of the ratings. Groups are defined by age-occupation.

MovieLens.

F. The PULL-ARM routine

The PULL-ARM routine used in Alg. 1 is described in Alg. 2.

G. Proofs

G.1. Theoretical results

G.1.1. USEFUL LEMMAS

Recall that OCEF considers a single audited group g .

The algorithm relies on valid confidence intervals. As in (Jamieson et al., 2014), we use anytime bounds inspired by the law of the iterated algorithm (LIL), and a union bound.

We say that a random variable is σ -subgaussian if it is subgaussian with variance proxy σ^2 . Since we assume the rewards for each user are bounded, more precisely $r_t^m \in [0, 1]$, they are $\frac{1}{2}$ -subgaussian.

Throughout the paper, we assume that rewards for each user are independent conditionally to the arm played.

Lemma 3. Let $\delta \in (0, 1)$. Assume the rewards are σ -subgaussian.

Let $\omega \in (0, 1)$, $\theta = \log(1 + \omega) \left(\frac{\omega \delta}{2(2 + \omega)} \right)^{\frac{1}{1 + \omega}}$.

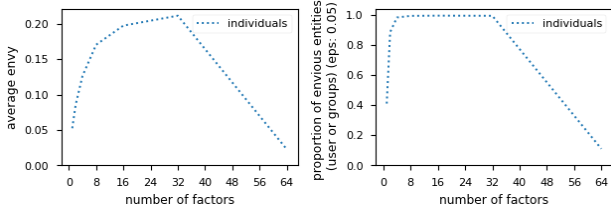


Figure 14. Model misspecification on Google Local with Softmax policy, trained on 20% of the ratings.

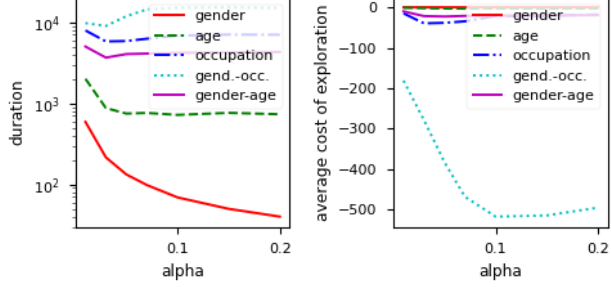


Figure 15. Scaling w.r.t. α on MovieLens experiments, with a sparsity level of 5%.

Let $r_t = \frac{1}{|g|} \sum_{m \in g} r_t^m$,

$$N_k(t) = \sum_{s=1}^t \mathbb{1}_{\{k_s=k\}} \quad \hat{\mu}_k(t) = \frac{\sum_{s=1}^t r_s \mathbb{1}_{\{k_s=k\}}}{N_k(t)}$$

$$\beta_k(t) = \sqrt{\frac{2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)}{|g|N_k(t)}} \times \sqrt{\log\left(\frac{2(K+1)}{\theta} \log((1 + \omega)N_k(t))\right)}$$

$$\underline{\mu}_k(t) = \hat{\mu}_k(t) - \beta_k(t) \quad \bar{\mu}_k(t) = \hat{\mu}_k(t) + \beta_k(t)$$

Then,

$$\mathbb{P}\left[\forall t > 0, \forall k \in [K], \mu_k \in [\underline{\mu}_k(t); \bar{\mu}_k(t)]\right] \geq 1 - \frac{\delta}{2}.$$

Importantly, $\beta_k(t)$ is in $\frac{1}{\sqrt{|g|}}$ because r_t is $\frac{\sigma}{\sqrt{|g|}}$ -subgaussian as an average of $|g|$ independent σ -subgaussian variables. Notice that the choice of θ makes sure that β_k is well defined as long as $N_k(t) > 0$. We use the convention that when $N_k(t) = 0$, $\beta_k(t)$ is strictly larger than when $N_k(t) = 1$ to ensure $\beta_g r_{arm}$ is strictly decreasing with N_k . Also, when $N_k(t) = 0$, we set $\hat{\mu}_k(t) = 0$.

Following (Garcelon et al., 2020), our lower bound on the conservative constraint relies on Freedman’s martingale inequality (Freedman, 1975).

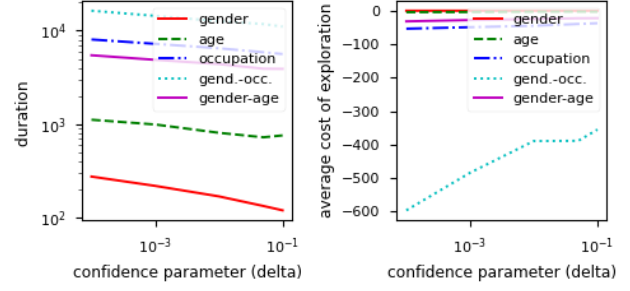


Figure 16. Scaling w.r.t. δ on MovieLens experiments, with a sparsity level of 5%.

Algorithm 2: PULL-ARM using transported policies (line 4).

input : arm k , timestep t , user group g
output : Observed group reward r_t

- 1 **foreach** user $m \in g$ **do** // in parallel
- 2 Observe context $x_t^m \sim q^m$;
- 3 **if** $k=0$ **then** $n = m$ // pull baseline;
- 4 **else** Sample “other user” $n \sim \tilde{\gamma}_{m,\cdot}^*$ // Def. 4.2;
- 5 Show $a_t^m \sim \pi^n(\cdot | x_t^m)$ and observe $r_t^m \sim \nu^m(a_t^m | x_t^m)$;
- 6 **end**
- 7 **return** $r_t = \frac{1}{|g|} \sum_{m \in g} r_t^m$;

Lemma. Assume all rewards are σ -subgaussian. Let $A_t = \{s \leq t : k_s \neq 0\}$ be the number of times a non-baseline arm $k \neq 0$ has been pulled up to time t . Let $\phi(t) = \frac{\sigma}{\sqrt{|g|}} \sqrt{2|A_{t-1}| \log\left(\frac{6|A_{t-1}|^2}{\delta}\right) + \frac{2}{3} \log\left(\frac{6|A_{t-1}|^2}{\delta}\right)}$.

Then, $\forall \delta > 0$,

$$\mathbb{P}\left[\forall t > 0, \left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \phi(t)\right] \geq 1 - \frac{\delta}{2}.$$

Notice that the first term in $\phi(t)$ is in $\frac{1}{\sqrt{|g|}}$ because r_t is $\frac{\sigma}{\sqrt{|g|}}$ -subgaussian. As in Lemma 3, we use the convention $\phi(t) = 0$ when $|A_{t-1}| = 0$.

Lemma 4. Let $\delta \in (0, 1)$.

Let $\Phi(t) = \min\left(\sum_{k=1}^K \beta_k(t-1)N_k(t-1), \phi(t)\right)$, with $\phi(t)$ defined in Lemma 4. Let \mathcal{E} be the event under which all confidence intervals are valid, i.e.:

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \quad \text{with}$$

$$\mathcal{E}_1 = \{\forall k \in \{0, \dots, K\}, \forall t > 0, \mu_k(t) \in [\underline{\mu}_k(t); \bar{\mu}_k(t)]\}$$

$$\mathcal{E}_2 = \left\{\forall t > 0, \left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \Phi(t)\right\}.$$

Then $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$.

Proof. By Lemma 3, $\mathbb{P}[\mathcal{E}_1] \geq 1 - \frac{\delta}{2}$. By the lemma above, with probability $1 - \frac{\delta}{2}$, we have for all $t > 0$, $|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)| \leq \phi(t)$.

Then, notice that

$$\left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| = \left| \sum_{k=1}^K N_k(t-1)(\mu_k - \hat{\mu}_k(t-1)) \right|.$$

Hence under \mathcal{E}_1 we also have:

$$\left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| \leq \sum_{k=1}^K N_k(t-1)\beta_k(t-1).$$

Therefore,

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{E}_1 \cap \left\{ \left| \sum_{s \in A_{t-1}} (\mu_{k_s} - r_s) \right| \leq \phi(t) \right\},$$

and thus, by a union bound, we have: $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$. \square

G.1.2. THEOREMS

We now provide our complete theoretical guarantees for correctness (Theorem 5), duration (Theorem 6) and cost (Theorem 7), which we then prove in App. G.2 and G.3. From these results, we derive Theorem 1 in the main paper, which we prove in App. G.4.

Theorem 5 (Correctness). *With probability at least $1 - \delta$:*

1. *OCEF satisfies the safety constraint (7) at every time step,*
2. *if OCEF outputs ϵ -NO ENVY then g is not ϵ -envious, and if it outputs ENVY, then g is envious.*

We denote $\log^+(\cdot) = \max(1, \log(\cdot))$.

Theorem 6 (Duration). *Let $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$, $\delta \in (0, 1)$, $\theta = \log(2)\sqrt{\frac{\delta}{6}}$, and*

$$\forall k \neq 0, \quad H_k = 1 + \frac{64}{|g|\eta_k^2} \log \left(\frac{2(K+1) \log^+ \left(\frac{128(K+1)}{\theta|g|\eta_k^2} \right)}{\theta} \right),$$

$$H_0 = \max \left(\max_{k \in [K]} H_k, \frac{6K+2}{\alpha\mu_0} + \sum_{k=1}^K \frac{256 \log \left(\frac{2(K+1) \log(2H_k)}{\theta} \right)}{\alpha\mu_0|g|\eta_k} \right).$$

With probability at least $1 - \delta$, OCEF stops in at most τ steps, with

$$\tau \leq \sum_{k=0}^K H_k.$$

Finally, we define the *cost of exploration* as the potential reward lost because of exploration actions, in our case the cumulative reward lost, on average over users in the group:

$$C_t = t\mu_0 - \sum_{s=1}^t \mu_{k_s}. \quad (10)$$

In the worst case, the following bound holds:

Theorem 7 (Cost of exploration). *Under the assumptions and notation of Theorem 6, let τ be the time step where OCEF stops. With probability $1 - \delta$, we have:*

$$C_\tau \leq \sum_{k: \mu_k < \mu_0} (\mu_0 - \mu_k) H_k \quad (11)$$

Certification for all groups The audit of the full system consists in running OCEF for every group. Since we are making multiple tests, we need to use a tighter confidence parameter for each group so that the confidence intervals simultaneously hold for all groups.

Corollary 7.1 (Online certification). *With probability at least $1 - \delta$, running OCEF simultaneously for all G groups, each with confidence parameter $\delta' = \frac{\delta}{G}$, we have:*

1. *for all $i \in [G]$ OCEF satisfies the constraints (7),*
2. *all groups for which OCEF returns ϵ -NO ENVY are not ϵ -envious of any other users, and all groups for which OCEF returns ENVY are envious of another group.*
3. *For every group, the bounds on the duration of the experiment and the cost of exploration given by Theorems 6 and 7 (using δ/G instead of δ) are simultaneously valid.*

G.2. Proof of Theorem 5

Proof. We assume that event \mathcal{E} holds true. Then all confidence intervals are valid, i.e., for all $k = 0, \dots, K$, $\mu_k(t) \leq \bar{\mu}_k(t)$, and $\sum_{s \in A_{t-1}} \mu_{k_s} \geq \sum_{s \in A_{t-1}} r_s - \Phi(t)$.

Let Z_t be the safety budget, defined as $Z_t = \sum_{s=1}^t \mu_{k_s} - (1 - \alpha)\mu_0 t$, so that the conservative constraint (7) is equivalent to $\forall t, Z_t \geq 0$. We have $Z_t = \sum_{s \in A_{t-1}} \mu_{k_s} + \mu_{k_t} + (N_0(t-1) - (1 - \alpha)t)\mu_0$. Therefore, ξ_t (eq. (8)) is a lower bound on the safety budget Z_t if ℓ_t is played. By construction of the algorithm, the safety constraint (7) is immediately satisfied since a pull that could violate it is not permitted.

By the validity of confidence intervals under \mathcal{E} , if OCEF stops because of the first condition, then $\exists k, \mu_k > \mu_0$. Therefore 0 is not ϵ -envious of k and OCEF is correct.

If OCEF stops because of the second condition, i.e., $\forall k, \bar{\mu}_k(t) \leq \underline{\mu}_0(t) + \epsilon$, then $\forall k, \mu_k \leq \mu_0 + \epsilon$. Therefore 0 is not envious and OCEF is correct.

Since $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$, OCEF satisfies the safety constraint and is correct with probability $\geq 1 - \delta$.

□

G.3. Proofs of Theorem 6 and Theorem 7

Notation For conciseness, we use $\tilde{K} = K + 1$, and

$$\psi_k(t) = \frac{2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega) \log\left(\frac{2\tilde{K}}{\theta} \log((1 + \omega)N_k(t))\right)}{|g|},$$

$$\text{so that } \beta_k(t) = \sqrt{\frac{\psi_k(t)}{N_k(t)}}.$$

We shall also use $\Gamma_\omega = \frac{2\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)}{|g|}$. We use the convention $\psi_k(t) = 0$ when $N_k(t) = 0$, and set $\beta_k(t)$ to some value strictly larger than when $N_k(t) = 1$.

We remind that $\omega \in (0, 1)$, $\theta = \log(1 + \omega) \left(\frac{\omega\delta}{2(2 + \omega)}\right)^{\frac{1}{1 + \omega}}$ and $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$. We denote by $\eta_{\min} = \min_{k \in [K]} \eta_k$.

Finally, we notice that under event \mathcal{E} (as defined in Sec. G.1.1), we have for all $k \in \{0, \dots, K\}$ and all t :

$$\mu_k + 2\beta_k(t) \geq \bar{\mu}_k(t) \geq \mu_k \geq \underline{\mu}_k(t) \geq \mu_k - 2\beta_k(t). \quad (12)$$

Lemma 8. *Under event \mathcal{E} , for every $k \in [K]$, if k is pulled at round t , then $4\beta_k(t) \geq \eta_k$.*

Proof of Lemma 8. Since k is pulled at t , the two following inequalities hold:

$$\bar{\mu}_k(t - 1) > \underline{\mu}_0(t - 1) + \epsilon \quad (13)$$

$$\underline{\mu}_k(t - 1) \leq \bar{\mu}_0(t - 1) \quad (14)$$

We prove them by contradiction. If (13) does not hold, then k should be discarded from the active set at time $t - 1$, and therefore cannot be pulled at t . Likewise, if (14) does not hold, then the algorithm stops at $t - 1$, so k cannot be pulled at t .

Using (13) and (12), we have:

$$\mu_k + 2\beta_k(t - 1) \geq \bar{\mu}_k(t - 1) > \underline{\mu}_0(t - 1) + \epsilon \geq \mu_0 - 2\beta_0(t - 1) + \epsilon.$$

Since 0 was not pulled at time t , we also have $\beta_0(t - 1) \leq \beta_k(t - 1)$, hence $4\beta_k(t - 1) \geq \mu_0 + \epsilon - \mu_k$.

Using (14) and (12) we have $\mu_k - 2\beta_k(t) \leq \mu_0 + 2\beta_0(t)$ and since $\beta_0(t) \leq \beta_k(t)$, we obtain $4\beta_k(t - 1) \geq \mu_k - \mu_0$.

□

In the following lemma, we recall that we denote $\log^+(\cdot) = \max(1, \log(\cdot))$.

Lemma 9. *Under event \mathcal{E} , $\forall \tau > 0, \forall k \in [K]$, we have*

$$\begin{aligned} N_k(\tau) &\leq H_k \quad \text{with} \\ H_k &= 1 + \frac{32\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)}{|g|\eta_k^2} \times \\ &\quad \log\left(\frac{2(K + 1) \log^+\left(\frac{64(K + 1)\sigma^2(1 + \sqrt{\omega})^2(1 + \omega)^2}{\theta|g|\eta_k^2}\right)}{\theta}\right) \end{aligned}$$

Proof. Let $\tau > 0$, $k \in [K]$, and let $t \leq \tau$ be last time step before τ at which k was pulled. If such a t does not exist, then $N_k(\tau) = 0$ and the result holds. In all cases, we have $N_k(t) = N_k(\tau)$.

We consider $t > 0$ from now on.

By Lemma 8, we have $4\beta_k(t - 1) \geq \eta_k$, and thus $N_k(t - 1) \leq \frac{16\psi_k(t - 1)}{\eta_k^2}$, which writes, if $N_k(t) > 0$:

$$\begin{aligned} N_k(t - 1) &\leq \frac{16\psi_k(t - 1)}{\eta_k^2} \\ &\leq \frac{16\Gamma_\omega}{\eta_k^2} \log\left(\frac{2\tilde{K}}{\theta} \log((1 + \omega)N_k(t - 1))\right). \end{aligned} \quad (15)$$

Using $\frac{1}{t} \log\left(\frac{\log((1 + \omega)t)}{\Omega}\right) \geq c \Rightarrow t \leq \frac{1}{c} \log\left(\frac{\log((1 + \omega)/c\Omega)}{\Omega}\right)$ (see Equation (1) in (Jamieson et al., 2014)) with $\Omega = \frac{\theta}{2\tilde{K}}$ and $c = \frac{\eta_k^2}{16\Gamma_\omega}$, we obtain

$$N_k(t - 1) \leq \frac{16\Gamma_\omega}{\eta_k^2} \log\left(\frac{2\tilde{K}}{\theta} \log\left(\frac{(1 + \omega)32\tilde{K}\Gamma_\omega}{\theta\eta_k^2}\right)\right) \quad (16)$$

Since $N_k(t) = N_k(t - 1) + 1$, using \log^+ instead of \log inside to deal with the case $N_k(t - 1) = 0$ gives the desired result. □

Lemma 10. *Under event \mathcal{E} , at every time step τ , we have*

$$\begin{aligned} N_0(\tau) &\leq \max\left(\max_{k \in [K]} H_k, \frac{6K + 2}{\alpha\mu_0}\right) \\ &\quad + \sum_{k=1}^K \frac{64\sigma^2(1 + \sqrt{\omega})^2(1 + \omega) \log\left(\frac{2(K + 1) \log((1 + \omega)H_k)}{\theta}\right)}{\alpha\mu_0|g|\eta_k} \end{aligned}$$

Proof. Let $\tau > 0$ and $t \leq \tau$ the last time 0 was pulled before τ . We assume $t > 0$.

Case 1: 0 was pulled because $\beta_0(t-1) > \min_{k \in [K]} \beta_k(t-1)$.

Then $N_0(\tau) = N_0(t-1) + 1 \leq 1 + \max_{k \neq 0} N_k(t-1)$.

By lemma 8, we thus have $N_0(\tau) \leq \max_{k \in [K]} H_k$.

Case 2: 0 was pulled because $\xi_t < 0$. Here the proof follows similar steps as that of Theorem 5 in (Wu et al., 2016).

$$\sum_{s \in A_{t-1}} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(t-1) + (N_0(t-1) - (1-\alpha)t)\bar{\mu}_0(t-1) < 0$$

We drop $\underline{\mu}_{\ell_t}(t-1)$, replace t by $\sum_{k=0}^K N_k(t-1) + 1$ and rearrange terms to obtain:

$$\begin{aligned} \alpha N_0(t-1)\bar{\mu}_0(t-1) &\leq (1-\alpha)\bar{\mu}_0(t-1) \\ &+ (1-\alpha) \sum_{k=1}^K N_k(t-1)\bar{\mu}_0(t-1) - \sum_{s \in A_{t-1}} r_s + \Phi(t) \end{aligned} \quad (17)$$

Since we have $\beta_0(t-1) \leq \beta_k(t-1)$ (otherwise we would be in case 1), and $A_{t-1} = \sum_{k=1}^K N_k(t-1)$, we bound the the sum over arms in (17):

$$\begin{aligned} &\sum_{k=1}^K N_k(t-1)\bar{\mu}_0(t-1) \\ &\leq \sum_{k=1}^K N_k(t-1)(\mu_0 + 2\beta_0(t-1)) \\ &\leq \sum_{k=1}^K N_k(t-1)(\mu_0 + 2\beta_k(t-1)) \\ &= \sum_{s \in A_{t-1}} \mu_0 + \sum_{k=1}^K 2\beta_k(t-1)N_k(t-1). \end{aligned}$$

Using Lemma 4, we also bound $-\sum_{s \in A_{t-1}} r_s \geq \sum_{s \in A_{t-1}} \mu_s + \Phi(t)$ (under \mathcal{E}).

Plugging this into (17) gives:

$$\begin{aligned} \alpha N_0(t-1)\bar{\mu}_0(t-1) &\leq (1-\alpha)\bar{\mu}_0(t-1) \\ &+ 2(1-\alpha) \sum_{k=1}^K N_k(t-1)\beta_k(t-1) \\ &+ \sum_{s \in A_{t-1}} ((1-\alpha)\mu_0 - \mu_{k_s}) + 2\Phi(t). \end{aligned}$$

Recall that $\Phi(t) = \min(\sum_{k=1}^K N_k(t-1)\beta_k(t-1), \phi(t))$, and therefore $\Phi(t) \leq \sum_{k=1}^K N_k(t-1)\beta_k(t-1)$.

Using $\mu_0 - \mu_{k_s} \leq \eta_{k_s}$ and $\sum_{s \in A_{t-1}} \eta_{k_s} = \sum_{k=1}^K N_k(t-1)\eta_k$, we obtain:

$$\begin{aligned} \alpha N_0(t-1)\bar{\mu}_0(t-1) &\leq (1-\alpha)\bar{\mu}_0(t-1) \\ &+ \sum_{k=1}^K \left((\eta_k - \alpha\mu_0)N_k(t-1) \right. \\ &\quad \left. + 4\sqrt{\Psi_k(t-1)N_k(t-1)} \right). \end{aligned}$$

We bound $f_k := (\eta_k - \alpha\mu_0)N_k(t-1) + 4\sqrt{\Psi_k(t-1)N_k(t-1)}$.

Since (15) $N_k(t-1) \leq \frac{16\psi_k(t-1)}{\eta_k^2} + 1$, and $\eta_k - \alpha\mu_0 \leq \eta_k$, we have

$$f_k \leq \frac{16\psi_k(t-1)}{\eta_k} + \eta_k + 4\sqrt{\frac{16\psi_k(t-1)^2}{\eta_k^2} + \psi_k(t-1)}$$

Using $\sqrt{(\frac{x}{z})^2 + x} \leq \frac{x}{z} + \frac{z}{2}$ for $x \geq 0, z > 0$, with $x = 4\psi_k(t-1)$ and $z = \eta_k$, we obtain:

$$\begin{aligned} f_k &\leq \frac{16\psi_k(t-1)}{\eta_k} + \frac{16\psi_k(t-1)}{\eta_k} + 3\eta_k \\ &\leq \frac{32\psi_k(t-1)}{\eta_k} + 3\eta_k. \end{aligned} \quad (18)$$

Using $\psi_k(t-1) = \Gamma_\omega \log\left(\frac{2\tilde{K}}{\theta} \log((1+\omega)N_k(t-1))\right)$ if $N_k(t) > 0$ and $N_k(t-1) \leq H_k$ by Lemma 9, we obtain

$$f_k \leq \frac{32\Gamma_\omega}{\eta_k} \log\left(\frac{2\tilde{K}}{\theta} \log((1+\omega)H_k)\right) + 3\eta_k.$$

This bound is also valid when $N_k(t) > 0$.

Going back to (17), and since $\mu_0 \leq \bar{\mu}_0(t-1)$ under \mathcal{E} , we have (notice $\eta_k \leq 2$ since $\mu_k \in [0, 1]$ and $\epsilon \in [0, 1]$):

$$\begin{aligned} \alpha N_0(t-1)\mu_0 &\leq (1-\alpha)\bar{\mu}_0(t-1) + 6K \\ &+ \sum_{k=1}^K \frac{32\Gamma_\omega}{\eta_k} \log\left(\frac{2\tilde{K}}{\theta} \log((1+\omega)H_k)\right). \end{aligned} \quad (19)$$

To bound the first term of the right-hand side, let us first notice that the final result holds if $N_0(t-1) \leq \max_{k \in [K]} H_k$. So we can assume $N_0(t-1) > \max_{k \in [K]} H_k$ from now on. By the definition of the H_k s (see above (15)), this implies $N_0(t-1) > \frac{16\psi_0(t-1)}{\eta_{\min}^2}$, which in turn implies $4\beta_0(t-1) \leq \eta_{\min}$.

We thus use $\bar{\mu}_0(t-1) \leq \mu_0 + 2\beta_0(t-1) \leq \mu_0 + \frac{\eta_{\min}}{2} \leq 2$, which gives the final result.

The result directly follows from (19). \square

The proof of Theorem 6 follows from $\tau = \sum_{k=1}^K N_k(\tau) + N_0(\tau)$, by setting $\omega = 1$ for ease of reading, and $\sigma = \frac{1}{2}$ since Bernoulli variables are $\frac{1}{2}$ -subgaussian (using Hoeffding's inequality (Hoeffding, 1963)).

We prove Corollary 7.1 from Theorem 5 and Theorem 6.

We now prove Theorem 7:

Proof. Since playing the baseline is neutral in the cost of exploration, it can be re-written as:

$$C_\tau = \sum_{k=1}^K (\mu_0 - \mu_k) N_k(\tau) \leq \sum_{k: \mu_k < \mu_0} (\mu_0 - \mu_k) N_k(\tau),$$

where τ is the time the algorithm stops. Using Lemma 9 to upper bound $N_k(\tau)$, we obtain the result. \square

Corollary 7.1 simply follows from the fact that by applying each algorithm with confidence δ/G , the confidence intervals are then simultaneously valid for all groups with probability $1 - \delta$, so all the correctness/duration/cost proofs holds for all groups simultaneously with probability $1 - \delta$.

G.4. Proof of Theorem 1

Theorems 5, 6, and 7 are summarized in Theorem 1 in the main paper. We restate *with corrections* Theorem 1 and prove it below:

Theorem. Let $\epsilon \in (0, 1]$, $\alpha \in (0, 1]$, $\delta \in (0, \frac{1}{2})$ and

$$\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k) \text{ and } h_k = \max(1, \frac{1}{|g|\eta_k}).$$

Using $\mu, \bar{\mu}$ and Φ given in Lemmas 3 and 4, OCEF achieves the following guarantees with probability at least $1 - \delta$:

- OCEF is correct and satisfies the conservative constraint on the recommendation performance (7).

- The duration is in $O\left(\sum_{k=1}^K \frac{h_k \log\left(\frac{K \log\left(\frac{K h_k}{\delta \eta_k}\right)}{\delta}\right)}{\min(\alpha \mu_0, \eta_k)}\right)$.

- The cost is in $O\left(\sum_{k: \mu_k < \mu_0} \frac{(\mu_0 - \mu_k) h_k}{\eta_k} \log\left(\frac{K \log\left(\frac{K h_k}{\delta \eta_k}\right)}{\delta}\right)\right)$.

Proof. With $\delta \in (0, \frac{1}{2})$, let $\theta = \log(2) \sqrt{\frac{\delta}{6}}$. Then Theorems 6 and 7 hold for (δ, θ) .

Duration We first show that:

$$H_k = O\left(\frac{h_k}{\eta_k} \log\left(\frac{K h_k}{\delta \eta_k}\right)\right), \quad (20)$$

$$\log(H_k) = O\left(\log\left(\frac{K h_k}{\delta \eta_k}\right)\right). \quad (21)$$

Recall from Th. 6 that H_k is defined as:

$$H_k = 1 + \frac{64}{|g|\eta_k^2} \log\left(\frac{2(K+1) \log^+\left(\frac{256(K+1)}{\theta |g|\eta_k^2}\right)}{\theta}\right)$$

We replace the \log^+ term from Th. 6 by $\log\left(\frac{K h_k}{\delta \eta_k}\right) > 0$, because $\frac{K h_k}{\delta} \geq 3$ as soon as $K \geq 2$. We thus have

$$H_k = 1 + O\left(\frac{1}{|g|\eta_k^2} \underbrace{\log\left(\frac{K}{\delta} \log\left(\frac{K h_k}{\delta \eta_k}\right)\right)}_{=B}\right), \quad (22)$$

Using $\log(x) \leq x \Rightarrow x \log(x) \leq x^2$ for $x \geq 0$, and the fact that $\log\left(\frac{K h_k}{\delta \eta_k}\right) \geq 0$, we have:

$$B \leq \log\left(\frac{K h_k}{\delta \eta_k} \log\left(\frac{K h_k}{\delta \eta_k}\right)\right) \leq 2 \log\left(\frac{K h_k}{\delta \eta_k}\right).$$

Since $1 + \frac{1}{|g|\eta_k^2} \leq 2 \frac{h_k}{\eta_k}$, eq. (20) holds.

We now bound $\log(H_k)$:

$$\log(H_k) = O\left(\log\left(\frac{h_k}{\eta_k} \log\left(\frac{K h_k}{\delta \eta_k}\right)\right)\right) \quad (23)$$

$$= O\left(\log\left(\frac{K h_k}{\delta \eta_k} \log\left(\frac{K h_k}{\delta \eta_k}\right)\right)\right) \quad (24)$$

$$= O\left(\log\left(\frac{K h_k}{\delta \eta_k}\right)\right) \quad (25)$$

where the last line comes from $\frac{K h_k}{\delta \eta_k} \log\left(\frac{K h_k}{\delta \eta_k}\right) \leq \left(\frac{K h_k}{\delta \eta_k}\right)^2$.

Therefore, eq. (21) holds.

Now, let

$$\Gamma = \frac{6K+2}{\alpha \mu_0} + \sum_{k=1}^K \frac{128 \log\left(\frac{2(K+1) \log(2H_k)}{\theta}\right)}{\alpha \mu_0 |g|\eta_k},$$

so that $H_0 = \max(\max_{k \in [K]} H_k, \Gamma)$.

We have:

$$\Gamma = O\left(\frac{K}{\alpha \mu_0} + \sum_{k=1}^K \frac{h_k}{\alpha \mu_0} \log\left(\frac{K \log(H_k)}{\delta}\right)\right)$$

$$= O\left(\sum_{k=1}^K \frac{h_k}{\alpha \mu_0} \log\left(\frac{K \log(H_k)}{\delta}\right)\right)$$

$$= O\left(\sum_{k=1}^K \frac{h_k}{\alpha \mu_0} \log\left(\frac{K \log\left(\frac{K h_k}{\delta \eta_k}\right)}{\delta}\right)\right),$$

where the second equality is because $K = \sum_{k=1}^K 1 \leq \sum_{k=1}^K h_k$, and the last equality uses eq. (21). Combining this with eq. (20) we have:

$$H_0 = O\left(\sum_{k=1}^K \frac{h_k}{\min(\alpha\mu_0, \eta_k)} \log\left(\frac{K \log(\frac{Kh_k}{\delta\eta_k})}{\delta}\right)\right).$$

Using eq. (20) again to bound $\tau = H_0 + \sum_{k=1}^K H_k$, we get the desired bound for duration.

Cost For the cost, we remind the bound given in Th. 7:

$$\begin{aligned} C_\tau &\leq \sum_{k:\mu_k < \mu_0} (\mu_0 - \mu_k) H_k \\ &= O\left(\sum_{k:\mu_k < \mu_0} \frac{(\mu_0 - \mu_k)h_k}{\eta_k} \log\left(\frac{K}{\delta} \log\left(\frac{Kh_k}{\delta\eta_k}\right)\right)\right) \end{aligned} \quad (26)$$

using (22) and $1 + \frac{1}{|g|\eta_k^2} = O(\frac{h_k}{\eta_k})$.

□