Feature-based Uncertainty Model for School Choice

Yao Zhang
Kyushu University
Fukuoka, Japan
zhang@agent.inf.kyushu-u.ac.jp

Makoto Yokoo
Kyushu University
Fukuoka, Japan
yokoo@inf.kyushu-u.ac.jp

ABSTRACT

In this work, we consider a school choice scenario where a student does not exactly know which college is better for her. Although it is hard for a student to obtain an exact preference, she can usually compare specific features of colleges, such as reputation, location, and campus facilities. Motivated by this, we propose a feature-based uncertainty model for school choice where a student's preference is based on a linear combination of her utilities over different features, and the coefficients of the combination are treated as random variables. Our main goal is to achieve a higher probability of stability (ProS) and incentive compatibility (IC) for students. Unfortunately, these two goals are incompatible in general. We show that a student-proposing deferred acceptance (DA) that prioritizes colleges with higher expected ranking can achieve a worst-case approximation ratio of $(1/n)^n$ on ProS, while a DA with a carefully defined iterated comparison vector can guarantee the strongest achievable form of IC. Finally, we provide additional results for some specific restrictions on the model.

KEYWORDS

School Choice, Uncertain Preferences, Stability

ACM Reference Format:
Yao Zhang and Makoto Yokoo. 2026. Feature-based Uncertainty Model for School Choice. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/KGWC6399

1 INTRODUCTION

School choice, as a classical many-to-one matching problem, has been extensively studied for decades due to its wide range of applications [14]. The primary objective is typically to design a procedure that assigns students to colleges such that the resulting matching satisfies desirable properties such as efficiency, stability, and incentive compatibility [3]. Beyond the classical model, numerous extensions incorporating additional settings or constraints have been proposed to capture more realistic application scenarios—for example, constraints on college seats [7, 15, 19], or networked structure among students [11, 26].

In most of the literature on school choice theory, it is typically assumed that students have well-defined preferences over all colleges. In reality, however, it is often difficult for students to express clear and consistent preferences due to time constraints and uncertainty about the future [10, 17, 20]. To address this, some studies introduce simplified or restricted models of student preferences [12, 16]. Nevertheless, students are generally aware of which features of schools they value and can often compare schools based on specific attributes such as course suitability, academic reputation, employment prospects, or teaching quality [24]. The main challenge for a student lies in not knowing which of these attributes will ultimately prove most important to her before experiencing them firsthand.

Therefore, motivated by this observation, we propose a *feature-based uncertainty model* for school choice, in which each student has a well-defined utility function for every feature of a college. A student's overall preference over colleges is determined by a linear combination of these feature utilities, where the coefficients in the combination are treated as random variables. The probability distributions of these coefficients are assumed to be known to the students themselves. On the other hand, colleges are assumed to have complete and deterministic preferences over students, as in the classical model.

Under this model, our main objective is to design a matching algorithm that ensures both stability and incentive compatibility for students. Regarding stability, since students' preferences involve uncertainty, we consider the probability of stability (ProS) and aim to find a matching that maximizes this probability. We show that, in general, this problem is computationally NP-hard, and therefore we instead seek approximation algorithms. For incentive compatibility (IC), the same uncertainty requires us to evaluate the probability of improvement that a student can obtain by misreporting her information. The strongest form of IC would guarantee that no possible improvement can ever be achieved through misreporting, which is too strong to obtain meaningful algorithms. A milder version ensures that the probability of any improvement from misreporting is less than 1/2, though this condition is incompatible with any non-zero approximation of ProS. The weakest form only requires that certain improvements from misreporting are prevented. We examine a family of student-proposing deferred acceptance algorithms with various well-designed proposing orders to achieve different combinations of these properties, as summarized in Table 1.

The overall paper is organized as follows. Section 2 briefly reviews the most related literature. Section 3 introduces the model and summarizes the main results, and Sections 4 to 7 present all the theoretical results[1]. Section 8 includes a discussion of future work.

2 RELATED WORK

Our work is most related to a body of literature on stable matching with uncertain preferences. Aziz et al. [6] first introduced the idea of stable marriage with uncertain linear preference, proposing three models of uncertainty: the lottery model, the compact indifference

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/KGWC6399

[1] See the full version for missing proofs: https://arxiv.org/pdf/2602.12615

Table 1: Summary of Theoretical Results for Four Proposing Methods: Lexicographic Order of Comparison Vectors (LOCV), Lexicographic Order of Iterated Comparison Vectors (LOICV), Higher Expected Ranking First (HERF), and Higher Expected Utility First (HEUF). IC-C corresponds to incentive compatibility (IC) that prevents certain improvement by cheating, and IC-R corresponds to IC that prevents $1/2$ probability of improvement by cheating.

Methods	Worst-case Approx. of ProS	Incentive Compatibility
LOCV	0	IC-C
LOICV	0	IC-C (IC-R when $\|F\| = 2$)
HERF	$(1/n)^n$	IC-C
HEUF	0	IC-C; (when $\|F\| = 2$, IC-R iff the probabilities of the first feature's weight being larger or less than its expectation are equal)

model and the joint probability model. They examined the computational complexity of several problems related to stability probability and showed that finding a matching with the highest stability probability is generally intractable. Later, they further introduced a model with uncertain pairwise preferences, where preferred relations can be cyclic [5]. Alimudin and Ishida [4] also studied the stable marriage with uncertain preferences using a randomized approach. Random lottery tie-breakers have also been considered in school choice problems [2]. Abdulkadiroğlu et al. [1] showed how lottery tie-breaking implements a stratified randomized trial within a DA match. In our model, if we restrict all college capacities to be one, it naturally induces a new form of uncertainty model for the stable marriage problem. A detailed comparison between our model and existing models is provided in the full version. Similar ideas have also been extended to other domains, including resource assignment [13], house allocation [8], and committee voting [9].

3 THE MODEL

We consider a school choice scenario where a student's preference is determined by a set of measurable features, while the relative importance of these features is uncertain. We assume that every college is acceptable to each student, and vice versa. More concretely, an instance of school choice with feature-based uncertain preferences $\mathcal{I} = (N, M, X, \succ_M, F, U, \mu)$ consists of:

- N: the set of all students, and denote them by $s_1, s_2, \ldots, s_n$.
- M: the set of all colleges, and denote them by $c_1, c_2, \ldots, c_m$.
- X: m positive integers $x_{c_1}, x_{c_2}, \ldots, x_{c_m}$, which represent the capacities of all colleges, i.e., for any $1 \le j \le m$, college c_j can accept at most x_{c_j} students.
- $\succ_M$: the strict preferences of colleges over all students. Specifically, for a college $c \in M$, $s_i \succ_c s_j$ means c prefers s_i to s_j.
- F: the set of features. Each feature $f \in F$ represents an aspect that a student can evaluate for all colleges.
- U: the set of utility functions. Each utility function $u_s^f : M \mapsto [0, 1]$ represents how much a student s values each college with respect to feature $f \in F$. We assume that these evaluations are normalized to $[0, 1]$. This normalization does not affect generality, since the aggregated preference of student s assigns different weights to different features.
- μ: the set of probability distributions of weights for different features. For each student s, μ_s is a probability distribution over all possible weight vectors, representing her uncertainty about the exact preference over colleges. Specifically, a weight vector w_s is $|F|$-dimensional and satisfies $w_s^f \ge 0$ for any $f \in F$ and $\sum_{f \in F} w_s^f = 1$. An aggregated preference with w_s and $\{u_s^f\}_{f \in F}$ is defined by the weighted utility[2], i.e., $c_i \succeq_s^{w_s} c_j$ if and only if $\sum_{f \in F} w_s^f u_s^f(c_i) \ge \sum_{f \in F} w_s^f u_s^f(c_j)$. We assume that μ_s and $\mu_{s'}$ are independent for any distinct students $s, s' \in N$, and both the p.d.fs ($\Pr[w_s = w]$) and c.d.fs ($\Pr[\forall f, w_s^f \le w^f]$) can be computed in constant time.

A matching π specifies the pairs formed between students and colleges. For simplicity of notation, let $\pi(s)$ denote the college matched to a student $s \in N$, and let $\pi(s) = \text{null}$ if s is unmatched; let $\pi(c)$ denote the set of students matched to a college $c \in M$, and $\pi(c) = \emptyset$ if c is unmatched. A matching algorithm is a procedure that returns a matching for any given instance of a matching problem. We consider the following properties for matching algorithms.

Probability of Stability (ProS). Given a matching π and the corresponding instance, we can examine whether π is stable under any possible aggregated preferences of the students. Since an aggregated preference of a student may contain ties, we consider a weaker notion of stability where a block must incur a strict improvement.

Definition 3.1. Given aggregated preferences of all students $\{\succeq_s^{w_s}\}_{s \in N}$, a matching π is (weak) stable if there is no (strong) block (s, c), $s \in N$, $c \in M$, such that $c \succ_s^{w_s} \pi(s)$, and $|\pi(c)| < x_c$ or $\exists s' \in \pi(c)$ with $s \succ_c s'$.

Given an instance, the probability of stability of a matching π, denoted by $\text{ProS}(\pi; \mathcal{I})$, is defined as the probability that π is stable under aggregated preferences sampled[3], according to the weight distributions μ in $\mathcal{I}$. For each instance, a matching π^* is called an optimal matching if, for any other matching π', we have $\text{ProS}(\pi^*) \ge \text{ProS}(\pi')$. Note that for any aggregated preference, there exists at least one stable matching, implying $\text{ProS}(\pi^*) > 0$. Our goal is to find a matching algorithm that achieves a ProS value as close as possible to the optimal one for any instance. Therefore, for a matching algorithm, we consider the **worst-case approximation ratio** with respect to an optimal matching.

Definition 3.2. A matching algorithm Alg is α-optimal, if it satisfies $\inf_{\mathcal{I}} \frac{\text{ProS}(\text{Alg}(\mathcal{I}); \mathcal{I})}{\max_\pi \text{ProS}(\pi; \mathcal{I})} \ge \alpha$. $\text{Alg}(\mathcal{I})$ denotes the output matching.

[2]Note that we will have a tie when the equality holds.
[3]If there is no ambiguity, we also simplify it by omitting $\mathcal{I}$ as $\text{ProS}(\pi)$

Incentive Compatibility (IC). We consider incentive compatibility from the students' side. In traditional settings, it means that a student may strategically misreport her preferences over colleges. In our setting, a student s can misreport both her utility functions $\{u_s^f\}_{f\in F}$ and probability distribution μ_s. Intuitively, incentive compatibility requires that no student can obtain a better outcome by strategically misreporting her information. Note that in our model, whether an outcome is actually better may itself be uncertain for a student, since her report must be decided *ex ante*. Therefore, we introduce several levels of IC, and by definition, we have that IC-A implies IC-R and IC-R implies IC-C.

Definition 3.3. For any instance $\mathcal{I}$ and any student $s \in N$, for simplicity, let $\pi(s; u_s, \mu_s)$ denote the college assigned to s by the matching algorithm Alg when she reports truthfully, and let $\pi'(s; u'_s, \mu'_s)$ denote the college matched to her under a misreport (either may be null). Then, the algorithm Alg is said to be

(1) **incentive compatible with certainty (IC-C)** if s cannot obtain a college that is certainly better than her truthful assignment, i.e., $\Pr[\pi'(s; u'_s, \mu'_s) \succ_s \pi(s; u_s, \mu_s)] = 1$ cannot occur, for all u_s, μ_s, u'_s, μ'_s;
(2) **incentive compatible with rationality (IC-R)** if s cannot obtain a college that has a higher probability of being better than her truthful assignment, i.e., $\Pr[\pi'(s; u'_s, \mu'_s) \succ_s \pi(s; u_s, \mu_s)] > 1/2$ cannot occur, for all u_s, μ_s, u'_s, μ'_s;
(3) **incentive compatible with adventurism (IC-A)** if s cannot obtain a college that has any chance of being better than her truthful assignment, i.e., $\Pr[\pi'(s; u'_s, \mu'_s) \succ_s \pi(s; u_s, \mu_s)] > 0$ cannot occur, for all u_s, μ_s, u'_s, μ'_s.

4 IMPOSSIBILITIES AND INTRACTABILITY

In this section, we present several impossibility and intractability results when seeking a matching with a higher probability of stability. All these negative results occur even when the number of features is $|F| = 2$. We begin by stating a useful lemma that provides a formula for calculating the pairwise probability that one college is preferred over another by a student.

Lemma 4.1. *If the number of features equals to 2, i.e., $F = \{f_1, f_2\}$, then for any student $s \in N$ and any two colleges $c_1, c_2 \in M$, denote $\Delta_s^f(c_i, c_j) = |u_s^f(c_i) - u_s^f(c_j)|$ for any $f \in F$, and $\eta_s(c_i, c_j) = 1 \Big/ \left(1 + \frac{\Delta_s^{f_1}(c_i,c_j)}{\Delta_s^{f_2}(c_i,c_j)}\right)$ (define $\eta_s(c_i, c_j) = 0$ if $\Delta_s^{f_2}(c_i, c_j) = 0$). We have*

$$\Pr[c_i \succ_s^{w_s} c_j] = \begin{cases} 1 & \text{if } u_s^{f_1}(c_i) > u_s^{f_1}(c_j) \text{ and } u_s^{f_2}(c_i) > u_s^{f_2}(c_j); \\ 0 & \text{if } u_s^{f_1}(c_i) \le u_s^{f_1}(c_j) \text{ and } u_s^{f_2}(c_i) \le u_s^{f_2}(c_j); \\ \Pr[w_s^{f_1} > \eta_s(c_i, c_j)] & \text{if } u_s^{f_1}(c_i) > u_s^{f_1}(c_j) \text{ and } u_s^{f_2}(c_i) \le u_s^{f_2}(c_j); \\ \Pr[w_s^{f_1} < \eta_s(c_i, c_j)] & \text{if } u_s^{f_1}(c_i) \le u_s^{f_1}(c_j) \text{ and } u_s^{f_2}(c_i) > u_s^{f_2}(c_j), \end{cases}$$

and the case of $\Pr[c_i \succeq_s^{w_s} c_j]$ can be generated by $1 - \Pr[c_j \succ_s^{w_s} c_i]$.

We observe that, when $|F| = 2$, the probability that one college is preferred over another is equivalent to the probability that the weight of the first feature lies within a certain interval. Based on this observation, we can derive an efficient method to calculate the probability of stability for any given matching.

Proposition 4.2. *Given an instance $\mathcal{I}$ with $|F| = 2$ and a corresponding matching π, we can compute the probability of stability $\mathrm{ProS}(\pi; \mathcal{I})$ in polynomial time.*

Proof. Since each probability distribution of a student's weight vector is independent from others, the probability of stability in general can be formulated as

$$\mathrm{ProS}(\pi; \mathcal{I}) = \prod_{s\in N} (1 - \Pr[\exists c \in M, s.t.\ (s, c) \text{ will be a block}]).$$

Consider each term in the product. For each student $s \in N$, define the potential set $\hat{C}_s(\pi) = \{c \neq \pi(s) \mid |\pi(c)| < x_c \text{ or } \exists s' \in \pi(c) \text{ with } s \succ_c s', \text{ and } \Pr[c \succ_s^{w_s} \pi(s)] > 0\}$. Then, only colleges in the potential set may have a positive probability to make a block with s. If $\hat{C}_s(\pi) = \emptyset$, then the term for s equals to 1. If there exists a college $c \in \hat{C}_s(\pi)$, such that $\Pr[c \succ_s^{w_s} \pi(s)] = 1$, then the term equals to 0 (and the whole ProS of π is 0). For the remaining cases, according to Lemma 4.1, the colleges in $\hat{C}_s(\pi)$ must belong to one of the following two categories:

- those c such that $u_s^{f_1}(c) > u_s^{f_1}(\pi(s))$ and $u_s^{f_2}(c) \le u_s^{f_2}(\pi(s))$. Denote the set of this category as $\hat{C}_s^1(\pi)$. For any $c \in \hat{C}_s^1(\pi)$, (s, c) will be a block when $w_s^{f_1} > \eta_s(c, \pi(s))$. Denote $\eta_s^1(\pi) = \min_{c\in\hat{C}_s^1(\pi)} \eta_s(c, \pi(s))$, and let $\eta_s^1(\pi) = 1$ if $\hat{C}_s^1(\pi) = \emptyset$. Then, $\Pr[\exists c \in \hat{C}_s^1(\pi), s.t.\ (s, c) \text{ is a block}] = \Pr[w_s^{f_1} > \eta_s^1(\pi)]$;
- those c such that $u_s^{f_1}(c) \le u_s^{f_1}(\pi(s))$ and $u_s^{f_2}(c) > u_s^{f_2}(\pi(s))$. Similarly, denote the set of this category as $\hat{C}_s^2(\pi)$. For any $c \in \hat{C}_s^2(\pi)$, (s, c) will be a block when $w_s^{f_1} < \eta_s(c, \pi(s))$. Denote $\eta_s^2(\pi) = \max_{c\in\hat{C}_s^2(\pi)} \eta_s(c, \pi(s))$, and let $\eta_s^2(\pi) = 0$ if $\hat{C}_s^2(\pi) = \emptyset$. Then, $\Pr[\exists c \in \hat{C}_s^2(\pi), s.t.\ (s, c) \text{ is a block}] = \Pr[w_s^{f_1} < \eta_s^2(\pi)]$.

Hence, the probability of whether there exists a block associated with s is the probability of whether the weight of the first feature $w_s^{f_1}$ lies in a specific interval, i.e.,

$$1 - \Pr[\exists c \in M, s.t.\ (s, c) \text{ will be a block}] = \begin{cases} 0 & \text{if } \eta_s^1(\pi) < \eta_s^2(\pi); \\ \Pr[\eta_s^2(\pi) \le w_s^{f_1} \le \eta_s^1(\pi)] & \text{if } \eta_s^1(\pi) \ge \eta_s^2(\pi), \end{cases}$$

The probability in the second case can be easily achieved by an oracle that can return the value of μ_s's c.d.f. Deriving the potential set, checking each college in it and computing the values of $\eta_s^1(\pi)$ and $\eta_s^2(\pi)$ totally need $O(m)$ time. Therefore, we can finally compute $\mathrm{ProS}(\pi)$ in $O(mn)$ time, which is polynomial. □

However, finding a matching with the highest probability of stability is NP-hard.

Theorem 4.3. *It is NP-hard to find a matching with the highest* ProS *in a school choice with feature-based uncertainty, even with the constraint of $|F| = 2$.*

Hence, we aim to find an algorithm that achieves a good approximation of the optimal ProS in general. Since we also require incentive compatibility for students, we examine what can be achieved under different levels of IC. The first impossibility result is the unattainability of a meaningful IC-A algorithm.

Proposition 4.4. *There is no matching algorithm that can satisfy the property of IC-A unless it always outputs the same result.*

Since IC-A is too strong to yield meaningful algorithms, we consider milder versions of IC. However, even IC-R is not compatible with achieving a non-zero approximation of ProS.

Theorem 4.5. *There is no matching algorithm that can be IC-R and α-optimal with $\alpha > 0$ even when the number of features $|F| = 2$.*

Proof. We can prove the statement by giving an instance with three students and three colleges. The utility functions are:

- s_1: $u_{s_1}^{f_1}(c_1) = 1.5\delta + 3\epsilon$, $u_{s_1}^{f_1}(c_2) = \delta + 2\epsilon$, $u_{s_1}^{f_1}(c_3) = 0$;
  $u_{s_1}^{f_2}(c_1) = 0$, $u_{s_1}^{f_2}(c_2) = 0.5\delta + 2\epsilon$, $u_{s_1}^{f_2}(c_3) = 1.5\delta + 2\epsilon$;
- s_2: $u_{s_2}^{f_{1/2}}(c_1) = 0.3$, $u_{s_2}^{f_{1/2}}(c_2) = 0.6$, $u_{s_2}^{f_{1/2}}(c_3) = 0.1$;
- s_3: $u_{s_3}^{f_{1/2}}(c_1) = 0.3$, $u_{s_3}^{f_{1/2}}(c_2) = 0.1$, $u_{s_3}^{f_{1/2}}(c_3) = 0.6$.

Here, δ and ϵ are any positive values, and $\epsilon \to 0$. We assume that the probability distributions of weights for all three students are uniform distributions over all possible weights. We can notice that, for students s_2 and s_3, whichever the weight vector is chosen, their aggregated preferences on colleges are always $c_2 \succ_{s_2} c_1 \succ_{s_2} c_3$ and $c_3 \succ_{s_3} c_1 \succ_{s_3} c_2$. On the other hand, the preferences of three colleges are $s_1 \succ_{c_1/c_2} s_2 \succ_{c_1/c_2} s_3$, and $s_1 \succ_{c_3} s_3 \succ_{c_3} s_2$. Here, we suppose the capacities of all colleges are one. First, we check what will happen if a matching algorithm matches s_1 to c_2. In this case, for the remaining students and colleges, we can only match s_2 to c_1 and s_3 to c_3; otherwise (s_2, c_1) will be a block. Since both c_1 and c_3 ranks s_1 the first in their preferences, the ProS of this matching is

$$\Pr[c_2 \succeq_{s_1}^{w_{s_1}} c_1 \text{ and } c_2 \succeq_{s_1}^{w_{s_1}} c_3] = \frac{\epsilon(\delta + 2\epsilon)}{(\delta + \epsilon)(\delta + 3\epsilon)},$$

which approaches 0 when $\epsilon \to 0$. Nevertheless, if we consider an alternative matching, where s_1 is matched to c_1, s_2 is matched to c, and s_3 is matched to c, then the ProS of this new matching is

$$\Pr[c_1 \succeq_{s_1}^{w_{s_1}} c_2 \text{ and } c_1 \succeq_{s_1}^{w_{s_1}} c_3] = (\delta + 2\epsilon)/(2\delta + 6\epsilon),$$

which does not approach 0 when $\epsilon \to 0$. Therefore, a matching algorithm that has a non-zero approximation ratio cannot match s_1 to c_2. However, for s_1, c_2 is more likely to be better than c_1 or c_3:

$$\Pr[c_2 \succeq_{s_1}^{w_{s_1}} c_1] = \frac{\delta + 4\epsilon}{2\delta + 6\epsilon} > \frac{1}{2}; \quad \Pr[c_2 \succeq_{s_1}^{w_{s_1}} c_3] = \frac{\delta + 2\epsilon}{2\delta + 2\epsilon} > \frac{1}{2}.$$

Then, whichever college the matching algorithm matches to s_1, she can construct a misreporting as follows.

- s_1: $u_{s_1}^{f_{1/2}}(c_1) = 0.3$, $u_{s_1}^{f_{1/2}}(c_2) = 0.6$, $u_{s_1}^{f_{1/2}}(c_3) = 0.1$,

where it is always $c_2 \succ_{s_1} c_1 \succ_{s_1} c_3$. In this new case, s_1 must be matched to c_2; otherwise (s_1, c_2) will be a block. Namely, s_1 can force a choice that will be better with probability larger than 1/2, which violates the property of IC-R. In summary, in the given instance, c_1 being matched to c_2 violates IC-R, while c_1 not being matched to c_2 violates the non-zero approximation ratio, which infers that a matching algorithm cannot be IC-R and α-optimal with $\alpha > 0$ at the same time. Since the instance only has two features, this impossibility holds even for the restriction of $|F| = 2$. □

Hence, we can only require IC-C if we want an algorithm with a positive approximation ratio. The final result in this section provides an upper bound for the ratio in this case.

Theorem 4.6. *There is no matching algorithm that can be IC-C and α-optimal with $\alpha > (\sqrt[3]{\phi})^n$, $\phi = \frac{\sqrt{5}-1}{2}$, even when $|F| = 2$.*

Proof. Since the approximation ratio is defined in the worst case, to establish this upper bound, it suffices to show that the ratio cannot be achieved in at least one instance. Consider the following case with $n = 3k$ students and $m = n$ colleges. Each college $c_j \in M$ has capacity $x_j = 1$. For each $1 \le i \le k$, we suppose that $s_{i+1} \succ_{c_i} s_i \succ_{c_i} s_{i+2} \succ_{c_i} \cdots$, $s_i \succ_{c_{i+1}} s_{i+1} \succ_{c_{i+1}} s_{i+2} \succ_{c_{i+1}} \cdots$, and $s_i \succ_{c_{i+2}} s_{i+2} \succ_{c_{i+2}} \cdots$, where "$\cdots$" means the preferences over all other students could be arbitrary. On the other hand, for the student side, we suppose the probability distributions of weights are uniform distributions, and

- s_i: $u_{s_i}^{f_1}(c_i) = 1 > u_{s_i}^{f_1}(c_{i+1}) = 0.8 > u_{s_i}^{f_1}(c_{i+2}) = 0.8 - y > \cdots$,
  $u_{s_i}^{f_2}(c_i) = 1 > u_{s_i}^{f_2}(c_{i+2}) = 0.8 > u_{s_i}^{f_2}(c_{i+1}) = y - 0.2 > \cdots$;
- s_{i+1}: $u_{s_{i+1}}^{f_1}(c_{i+1}) = 1 > u_{s_{i+1}}^{f_1}(c_i) = 1 - z > \cdots$,
  $u_{s_{i+1}}^{f_2}(c_i) = 1 > u_{s_{i+1}}^{f_2}(c_{i+1}) = z > \cdots$;
- s_{i+2}: $u_{s_{i+2}}^{f_1}(c_{i+2}) = 1 > \cdots$, $u_{s_{i+2}}^{f_2}(c_{i+2}) = 1 > \cdots$,

where "$\cdots$" means the utilities of all other colleges are arbitrary but with the same order in both f_1 and f_2. Then, the only uncertainties in preferences of these three students are $\Pr[c_{i+1} \succ_{s_i}^{w_{s_i}} c_{i+2}] = y$ and $\Pr[c_{i+1} \succ_{s_{i+1}}^{w_{s_{i+1}}} c_i] = z$.

Now, to have a matching with positive ProS, students s_i, s_{i+1} and s_{i+2} must be matched to colleges c_i, c_{i+1} and c_{i+2}: if s_i has not been matched to one of them, (s_i, c_{i+1}) would be a block with probability one; if s_{i+1} has not been matched to one of them, (s_{i+1}, c_i) would be a block with probability one; if s_{i+2} has not been matched to one of them, then at least one of three colleges could not enroll both s_i and s_{i+1}, which could make a block with s_{i+2} of probability one. Then, for all possible matching results among these six agents, there are only three of them that could have a positive ProS:
(a) $\{(s_i, c_i), (s_{i+1}, c_{i+1}), (s_{i+2}, c_{i+2})\}$ with ProS = z,
(b) $\{(s_i, c_{i+1}), (s_{i+1}, c_i), (s_{i+2}, c_{i+2})\}$ with ProS = y, and
(c) $\{(s_i, c_{i+2}), (s_{i+1}, c_i), (s_{i+2}, c_{i+1})\}$ with ProS = $(1 - z)(1 - y)$.
Consider the two extreme cases where $y = 0$ (Case 0) and $y = 1$ (Case 1). In Case 0, we can only choose matching (a) or (c) to ensure a positive ProS, while in Case 1, we can only choose matching (a) or (b). Notice that the student s_i can misreport her utilities to freely turn one case into the other. If in Case 0, s_i is matched to c_i but in Case 1, she is not matched to c_i, then an s_i with true type being Case 1 can misreport to pretend to be the one in Case 0, where she can be matched to her certainly best college c_i. Similarly, we cannot match s_i to c_i in Case 1, without matching her to c_i in Case 0. Hence, an IC-C matching algorithm with any positive approximation ratio can only have two choices: choosing matching (a) in both cases, or choosing (c) in Case 0 and choosing (b) in Case 1. For the former one, the approximation ratio of the best ProS among these agents is $\min\{z, \mathbb{I}(z \ge 1/2) + \mathbb{I}(z < 1/2) \cdot (z/(1 - z))\} = z$, where $\mathbb{I}(\cdot)$ is the indicator function; for the latter one, the approximation ratio of the best ProS among these agents is $\mathbb{I}(z \le 1/2) + \mathbb{I}(z > 1/2) \cdot ((1 - z)/z)$. Therefore, for any satisfiable matching algorithm, it will not exceed $\min_z \max\left\{z, \mathbb{I}(z \le 1/2) + \mathbb{I}(z > 1/2) \cdot \frac{1-z}{z}\right\} = \frac{\sqrt{5}-1}{2}$. Finally, since the above result is independent of i, an upper bound of the approximation ratio for the whole instance will be $\phi^k = \phi^{n/3}$, where $\phi = \frac{\sqrt{5}-1}{2}$. □

5 METHODS

In this section, we propose the methods that will be examined. Since we consider both IC and stability, we focus on a generalized version of the student-proposing deferred acceptance (GDA) algorithm [23], such that when the scenario degenerates to the case with no uncertainty among students, the result naturally coincides with the standard student-proposing DA algorithm.

Generalized Student-Proposing Deferred Acceptance

INPUT: an instance $(N, M, X, \succ_M, F, U, \mu)$.

(1) Initialize the matching π to be empty.
(2) Initialize the set of unmatched students as $N_{\mathcal{U}} \leftarrow N$.
(3) For each student $s \in N$, initialize the set of colleges which have rejected her as $R_s \leftarrow \emptyset$.
(4) **while** $|N_{\mathcal{U}}| > 0$ **and** for any $s \in N_{\mathcal{U}}$, $R_s \neq M$ **do**
  (a) Initialize P_c to be empty for all $c \in M$
  (b) For each student $s \in N_{\mathcal{U}}$ with $R_s \neq M$, let s propose to a college $c = \mathsf{Next}(s, R_s, \{u_s^f\}_{f\in F}, \mu_s)$. Add s to P_c.
  (c) For each college c such that $P_c \neq \emptyset$,
    if $|\pi(c)| + |P_c| \leq x_c$ **then**
      Add all students in P_c to $\pi(c)$.
      For each student $s \in P_c$, update $\pi(s) = c$ and remove s from $N_{\mathcal{U}}$.
    else
      Choose the most preferred x_c students of c from $\pi_c \cup P_c$, update $\pi(c)$ as the set of these x_c students and reject others.
      For each student s in updated $\pi(c)$, update $\pi(s) = c$ and remove s from $N_{\mathcal{U}}$ (if $s \in N_{\mathcal{U}}$).
      For each student s being rejected, update $\pi(s) =$ null and add s to $N_{\mathcal{U}}$ (if $s \notin N_{\mathcal{U}}$).
    end if
  end while

OUTPUT: the final matching π.

For each student s, we can define an ordering of colleges based on the Next() function. Denote this ordering as $\succ_s^{\mathsf{Next}}$. Then, the output of the algorithm must be student-optimal with respect to this evaluation. Specifically, among all possible matching results in which no student s can find a college that is more preferred under $\succ_s^{\mathsf{Next}}$ than her current assignment and that also prefers her to at least one of its current enrollees, each student receives her best possible option according to $\succ_s^{\mathsf{Next}}$. The focus on such student-optimal matching algorithms (e.g., those that are student-optimal in terms of expected utilities) is another reason we restrict attention to this class of methods.

Within the class of GDA, the key problem is designing a suitable Next() function. One of the most straightforward ways is described above: it selects the college with the highest expected utility among those not in R_s. Besides, recall that our main goal is seeking high ProS and IC, which may not be fully captured by expected utilities. Therefore, we propose three additional methods based on the probabilities of one college being preferred over another. More concretely, the four methods are defined as follows.

Higher Expected Utility First (HEUF). In the method of HEUF, the Next() function just chooses a college such that

$$\mathsf{Next}(s, R_s, \{u_s^f\}_{f\in F}, \mu_s) \in \arg\max_{c\notin R_s} \mathbb{E}_{w_s\sim\mu_s}\left[\sum_{f\in F} w_s^f \cdot u_s^f(c)\right],$$

with random or any predetermined tie-breaking. The time complexity of computing the expected utility depends on the forms of probability distributions. In real applications, it is usually a discrete distribution with finite support or a common parameterized continuous distribution, which can be computed efficiently.

Lexicographic Order of Comparison Vectors (LOCV). In the method of LOCV, the Next() function chooses a college based on a specific order of the colleges. For each student s, such an order is defined by a lexicographic order of *comparison vectors*. Given a college c, the comparison vector for the student s is defined as

$$q_s(c) = \mathsf{Ascending}((\Pr[c \succeq_s^{w_s} c'])_{c\in M, c'\neq c}),$$

where Ascending$(\cdot)$ rearranges a vector's elements in an ascending order. Then, for any two colleges c and c', we say $c \succ_s^{\mathsf{lex}} c'$, if there exists an integer $0 \leq k < m$, the first k elements in $q_s(c)$ and $q_s(c')$ is the same, while the $(k+1)$th element of $q_s(c)$ is larger than that of $q_s(c')$. Based on $\succ_s^{\mathsf{lex}}$, we can have a complete order of all colleges with random or any predetermined tie-breaking, and each time the Next() function returns the first college that has not rejected the student in $\succ_s^{\mathsf{lex}}$. The intuition behind LOCV is that we try to choose a college that has a higher chance of being better than another college in the worst case. By Lemma 4.1, these probabilities needed in the method can be efficiently computed by an oracle of c.d.f when $|F| = 2$. For $|F| \geq 3$, we discuss the issue in Section 7.2.

Lexicographic Order of Iterated Comparison Vectors (LOICV). The only difference compared to the LOCV method is that we update the comparison vectors iteratively based on the current R_s set:

$$\tilde{q}_s(c, R_s) = \mathsf{Ascending}((\Pr[c \succeq_s^{w_s} c'])_{c\in M\setminus R_s, c'\neq c}).$$

Then, each time, we can have an updated complete order for remaining colleges that have not been proposed by s, and the Next() function returns the first college in such an order. The intuition behind LOICV is that we do not need to worry about colleges that have rejected s making a block with her. The time complexity of the LOICV method is quadratic in that of the LOCV method, due to its additional computed probabilities.

Higher Expected Ranking First (HERF). In the method of HERF, we also consider the order of colleges iteratively as in LOICV. The idea is motivated by the observation that $\Pr[c_i \succ_s^{w_s} c_j] \geq \Pr[c_j \succ_s^{w_s} c_i]$ for any $j \neq i$ still does not imply c_i has the highest probability to be most preferred by the student s (see the third instance in Example 5.1). Hence, in HERF, the Next() function instead chooses a college such that

$$\mathsf{Next}(s, R_s, \{u_s^f\}_{f\in F}, \mu_s) \in \arg\max_{c\notin R_s} \Pr\left[\bigcap_{c'\neq c, c\notin R_s} c \succeq_s^{w_s} c'\right],$$

with random or any predetermined tie-breaking. By Lemma 4.1, the above probability is equal to the probability that the weight of the first feature locates in an intersection of $m-1$ intervals when $|F| = 2$, which can be computed efficiently by an oracle of c.d.f. For $|F| \geq 3$, we discuss the issue in Section 7.2.

With the above methods, we can obtain student-optimal matching results with respect to expected utilities, pairwise preferences, or expected rankings, and they all degenerate to standard student-proposing DA when uncertainty disappears. In the following example, we will see that, for the objective of ProS, none of these methods consistently outperforms the others across all instances. This suggests that each method may have its own suitable scenarios in practical applications.

Example 5.1. We will show three instances in this example. All of them have $n = m = 3$, $|F| = 2$, each college c has capacity $x_c = 1$, and each probability distribution μ_s is a uniform distribution. For the first instance, the preferences of colleges are $s_1 \succ_{c_1} s_2 \succ_{c_1} s_3$, $s_1 \succ_{c_2} s_3 \succ_{c_2} s_2$, and $s_2 \succ_{c_3} s_3 \succ_{c_3} s_1$. On the other hand, the utility functions of students are

- s_1: $u_{s_1}^{f_1}(c_1) = 0.3,\ u_{s_1}^{f_1}(c_2) = 0.2,\ u_{s_1}^{f_1}(c_3) = 1.0;$
  $u_{s_1}^{f_2}(c_1) = 0.7,\ u_{s_1}^{f_2}(c_2) = 0.4,\ u_{s_1}^{f_2}(c_3) = 0.3;$
- s_2: $u_{s_2}^{f_1}(c_1) = 0.5,\ u_{s_2}^{f_1}(c_2) = 0.1,\ u_{s_2}^{f_1}(c_3) = 0.7;$
  $u_{s_2}^{f_2}(c_1) = 0.6,\ u_{s_2}^{f_2}(c_2) = 0.3,\ u_{s_2}^{f_2}(c_3) = 0.1;$
- s_3: $u_{s_3}^{f_1}(c_1) = 0.9,\ u_{s_3}^{f_1}(c_2) = 0.3,\ u_{s_3}^{f_1}(c_3) = 0.6;$
  $u_{s_3}^{f_2}(c_1) = 0.2,\ u_{s_3}^{f_2}(c_2) = 0.3,\ u_{s_3}^{f_2}(c_3) = 0.1.$

Taking student s_3 as an example, the comparison vector $q_{s_3}(c_2) = (1/7, 2/5)$ since we can compute that $\Pr[c_2 \succeq_{s_3}^{w_{s_3}} c_1] = 1/7 < \Pr[c_2 \succeq_{s_3}^{w_{s_3}} c_3] = 2/5$. Similarly, we have $q_{s_3}(c_1) = (6/7, 1)$ and $q_{s_3}(c_3) = (0, 3/5)$. Hence, $c_1 \succ_{s_3}^{\text{lex}} c_2 \succ_{s_3}^{\text{lex}} c_3$. If applying the LOCV method, s_2 and s_3 will apply to c_1, and s_1 will apply to c_3 in the first round. c_1 then rejects s_3, and s_3 will then apply to c_2. The final matching is $\{c_1 : s_2, c_2 : s_3, c_3 : s_1\}$, with pairs (s_1, c_1), (s_1, c_2), (s_2, c_3) and (s_3, c_3) having positive probability to be blocks. Then, the corresponding ProS equals to $2/11$.

On the contrary, if applying the LOICV method, when c_1 rejects s_3 in the first round, s_3 will then apply to c_3 instead because $\tilde{q}_{s_3}(c_2, \{c_1\}) = (2/5)$ and $\tilde{q}_{s_3}(c_3, \{c_1\}) = (3/5)$. c_3 then has choice to reject s_1, c_1 then accepts s_1, and c_3 can get its best choice s_2. The final matching is $\{c_1 : s_1, c_2 : s_3, c_3 : s_2\}$, with no pairs having positive probability to be blocks. Then, the ProS equals to 1. We can observe that, compared to the LOCV method, the outcome for s_3, who actually has a different proposing order, does not change. Lastly, if applying the HEUF method, the proposing orders of the students are the same as those in the LOICV method, since $\mathbb{E}[u_{s_3}(c_1)] = 0.55 > \mathbb{E}[u_{s_3}(c_3)] = 0.35 > \mathbb{E}[u_{s_3}(c_2)] = 0.3$. Hence, the matching result of the HEUF method is the same as that of the LOICV method. Similarly, if applying the HERF method, the proposing orders of the students are also the same, which leads to the same result.

We can observe that in the first instance, applying the LOICV, HEUF, or HERF method is the best choice for a higher ProS. Next, we will give the second instance where the LOCV method becomes better than LOICV, which suggests iteratively updating the comparison vector is not always a better choice. The preferences of colleges in the second instance are $s_3 \succ_{c_1} s_1 \succ_{c_1} s_2$ and $s_2 \succ_{c_2/c_3} s_3 \succ_{c_2/c_3} s_1$, and the utility functions of students are

- s_1: $u_{s_1}^{f_1}(c_1) = 0.9,\ u_{s_1}^{f_1}(c_2) = 0.7,\ u_{s_1}^{f_1}(c_3) = 0.75;$
  $u_{s_1}^{f_2}(c_1) = 0.1,\ u_{s_1}^{f_2}(c_2) = 0.7,\ u_{s_1}^{f_2}(c_3) = 0.8;$
- s_2: $u_{s_2}^{f_{1/2}}(c_1) = 0.9,\ u_{s_2}^{f_{1/2}}(c_2) = 0.5,\ u_{s_2}^{f_{1/2}}(c_3) = 0.1;$
- s_3: $u_{s_3}^{f_{1/2}}(c_1) = 0.5,\ u_{s_3}^{f_{1/2}}(c_2) = 0.1,\ u_{s_3}^{f_{1/2}}(c_3) = 0.9.$

Notice that for student s_2 and s_3, their preferences are actually certain because the utility functions for all features are the same. For student s_1, if applying the LOCV method, she will first propose to c_3, and then propose to c_1 if being rejected by c_3. The final matching is $\{c_1 : s_1, c_2 : s_2, c_3 : s_3\}$, with no pairs having positive probability to be blocks (ProS = 1). On the contrary, if applying the LOICV method, s_1 will propose to c_2 if being rejected by the first choice c_3. Then, the final matching is $\{c_1 : s_2, c_2 : s_1, c_3 : s_3\}$, with the pair (s_1, c_1) having probability of $1/4$ to be a block (ProS = $3/4$). Lastly, the HEUF and HERF share the same proposing order of students in the LOICV method, so they have the same result.

In the above two instances, the LOICV, HEUF, and HERF methods share the same results. For the HEUF, it always shares the same result with the LOICV as long as the probability distributions satisfy certain conditions (see Theorem 6.7). However, it is not true for the HERF method, as we can see in the third instance. For this last instance, the preferences of colleges are

- c_1/c_3: $s_1 \succ_{c_1/c_3} s_3 \succ_{c_1/c_3} s_2$; $\quad c_2$: $s_3 \succ_{c_2} s_2 \succ_{c_2} s_1$,

and the utility functions of students are

- $s_{1/2}$: $u_{s_{1/2}}^{f_1}(c_1) = 0.25,\ u_{s_{1/2}}^{f_1}(c_2) = 0.3,\ u_{s_{1/2}}^{f_1}(c_3) = 0.8;$
  $u_{s_{1/2}}^{f_2}(c_1) = 0.4,\ u_{s_{1/2}}^{f_2}(c_2) = 0.3,\ u_{s_{1/2}}^{f_2}(c_3) = 0.7;$
- s_3: $u_{s_3}^{f_1}(c_1) = 0.1,\ u_{s_3}^{f_1}(c_2) = 1.0,\ u_{s_3}^{f_1}(c_3) = 0.8;$
  $u_{s_3}^{f_2}(c_1) = 1.0,\ u_{s_3}^{f_2}(c_2) = 0.2,\ u_{s_3}^{f_2}(c_3) = 0.5.$

For students s_1 and s_2, proposing order is always c_3, c_1 and c_2 in all methods. For student s_3, she will first propose to c_3 with the LOCV or the LOICV method because both $\Pr[c_3 \succeq_{s_3}^{w_{s_3}} c_1] = 7/12$ and $\Pr[c_3 \succeq_{s_3}^{w_{s_3}} c_2] = 3/5$ is larger than $1/2$. If being rejected by c_3, s_3 will then propose to c_1 in the LOCV, and propose to c_2 in the LOICV. Hence, for LOCV, the final matching is $\{c_1 : s_3, c_2 : s_2, c_3 : s_1\}$, with the pair (s_3, c_2) having probability of $9/17$ to be a block (ProS = $8/17$), while for LOICV, the final matching is $\{c_1 : s_2, c_2 : s_3, c_3 : s_1\}$, with the pair (s_3, c_1) having probability of $8/17$ to be a block (ProS = $9/17$). The results of HEUF and LOICV are the same.

If applying the HERF method, s_3 will not first propose c_3 because $\Pr[c_3 \succeq_{s_3}^{w_{s_3}} c_1 \text{ and } c_3 \succeq_{s_3}^{w_{s_3}} c_2] = \Pr[5/12 \le w_{s_3} \le 3/5] = 11/60$, which suggests that c_3 has the smallest probability to be the favorite college of s_3. In this case, s_3 will first propose to c_1, and finally lead to the same matching result as the LOCV method.

From the three instances in this example, we can observe that none of the four methods dominates the others, i.e., for each method, there exist instances where it is not the best choice. Although in special cases we cannot say that one method is always better than another, one may still be curious about their average performance across a large number of instances. A simulation experiment is provided in the full version, which offers some insights on it.

6 PROPERTIES OF METHODS

In this section, we analyze the properties of these four methods.

6.1 Probability of Stability

For the probability of stability, only the method of HERF can have a positive approximation ratio.

Theorem 6.1. *For GDA with the HEUF, LOCV, and LOICV methods, the worst-case approximation ratios to the optimal probability of being stable are all 0 even when the number of features $|F| = 2$.*

Theorem 6.2. *The GDA with the HERF method is $\left(\frac{1}{n}\right)^n$-optimal.*

Proposition 6.3. *The worst-case approximation ratio characterized in Theorem 6.2 is tight even when the number of features $|F| = 2$.*

Proof. We illustrate the tightness of the approximation ratio by showing an instance with $|F| = 2$ that meets the worst case scenario. Consider an instance with n students and $m = n$ colleges, and each college has a capacity of one. For each college $c_j \in M$, suppose s_j is the least preferred student, and s_{j+1} (s_1 if $j = n$) is the most preferred student. On the other hand, for each student $c_i \in N$, the probability distribution μ_s is a uniform distribution over all possible weight values, and her utility function is as follows.

- $u_{s_i}^{f_1}(c_j) = v_j$ for all $j \neq i$, and $u_{s_i}^{f_1}(c_i) = v_i + \epsilon$;
- $u_{s_i}^{f_1}(c_j) = v_{n-j+1}$ for all $j \neq i$, and $u_{s_i}^{f_1}(c_i) = v_{n-i+1} + \epsilon$,

where v is defined as: (i) $v_1 = 0$, and (ii) $v_k = v_{k-1} + (n-k+1)\delta$ for $k > 1$. Here, $0 < \delta \leq \frac{2}{n(n-1)}$ and ϵ is a very small positive number. This utility function partitions the intervals of $w_{s_i}^{f_1}$ as follows.

- For $1 \leq j < i-1$, c_j will be ranked the first in the aggregated preference of s_i when $\frac{j-1}{n} \leq w_{s_i}^{f_1} \leq \frac{j}{n}$;
- For $j = i-1$, c_j will be ranked the first in the aggregated preference of s_i when $\frac{i-2}{n} \leq w_{s_i}^{f_1} \leq \frac{i-1}{n} - \frac{\epsilon}{n\delta}$;
- For $j = i$, c_j will be ranked the first in the aggregated preference of s_i when $\frac{i-1}{n} - \frac{\epsilon}{n\delta} \leq w_{s_i}^{f_1} \leq \frac{i}{n} + \frac{\epsilon}{n\delta}$;
- For $j = i+1$, c_j will be ranked the first in the aggregated preference of s_i when $\frac{i}{n} + \frac{\epsilon}{n\delta} \leq w_{s_i}^{f_1} \leq \frac{i+1}{n}$;
- For $i+1 < j \leq n$, c_j will be ranked the first in the aggregated preference of s_i when $\frac{j-1}{n} \leq w_{s_i}^{f_1} \leq \frac{j}{n}$.

Since $w_{s_i}^{f_1}$ is a uniformly distributed random variable over $[0,1]$, c_i then has the highest probability to be ranked first by s_i. Then, the final result π of the HERF method matches s_i to c_i for every i. However, since c_i likes s_i the least, then each pair (s_i, c_j) with $j \neq i$ is a potential block. The final probability of stability is

$$\text{ProS}(\pi) = \prod_{i=1}^{n} \left(\frac{i}{n} + \frac{\epsilon}{n\delta} - \left(\frac{i-1}{n} - \frac{\epsilon}{n\delta} \right) \right) = \left(\frac{1}{n} + \frac{2\epsilon}{n\delta} \right)^n,$$

which approaches to $(1/n)^n$ when $\epsilon \to 0$. On the other hand, consider an alternative matching π' such that s_i is matched to c_{i-1} (c_n if $i = 1$). Since each college gets its favorite student, then $\text{ProS}(\pi') = 1$. Therefore, the worst-case approximation ratio is no larger than $(1/n)^n$. □

6.2 Incentive Compatibility

First, for all four methods, they all satisfy the minimal requirement of incentive compatibility.

Theorem 6.4. *The GDA with the LOCV, LOICV, HERF, or HEUF method is IC-C.*

Additionally, for LOICV, it can further satisfy IC-R if $|F| = 2$.

Lemma 6.5. *When $|F| = 2$, then for any student $s \in N$, and any three different colleges c_i, c_j, $c_k \in M$, if $\Pr[c_i \succeq_s^{w_s} c_j] \geq \frac{1}{2}$ and $\Pr[c_j \succeq_s^{w_s} c_k] \geq \frac{1}{2}$, we must have $\Pr[c_i \succeq_s^{w_s} c_k] \geq \frac{1}{2}$.*

Theorem 6.6. *The GDA with the LOICV method is IC-R when the number of features $|F| = 2$.*

Proof. First, as stated in the proof of Theorem 6.4, for any $s \in N$, no misreport can make her be accepted by a college that rejected her when she reported truthfully, so we only need to consider colleges that have not been proposed to by s after the algorithm terminates.

By Lemma 6.5, we can make a corollary that for any student s, and any subset of colleges $M' \subseteq M$ with $|M'| > 1$, there must exist a college $c \in M'$, such that $\Pr[c \succeq_s^{w_s} c'] \geq \frac{1}{2}$, for any other $c' \in M'$. The statement is obviously true for $|M'| = 2$. For $|M'| > 2$, it can be shown by contradiction. Suppose that for any $c \in M'$, there is at least one another college $c' \in M'$ such that $\Pr[c' \succ_s^{w_s} c] > \frac{1}{2}$. Then, we can construct a directed graph where nodes are colleges in M', and an edge (c, c') appears if and only if $\Pr[c' \succ_s^{w_s} c] > \frac{1}{2}$. Since each node has at least one out-edge appoints to the other node, there must be at least one cycle in the graph. Denote the cycle as $(c_1, c_2, \ldots, c_k, c_1)$, and we have $\Pr[c_i \succ_s^{w_s} c_{i+1}] > \frac{1}{2}$ for $1 \leq i < k$. By Lemma 6.5, we must also have $\Pr[c_1 \succeq_s^{w_s} c_k] \geq \frac{1}{2}$, which contradicts to $\Pr[c_k \succ_s^{w_s} c_1] > \frac{1}{2}$.

Then, according to the process of the LOICV method, in each round, the college proposed by a student s must be one of the colleges described above, with M' as the set of colleges that have not rejected s, since the comparison vectors are in the ascending order of elements. Therefore, for any college c' that s can be matched with by misreporting, compared to the college c she can be matched without misreporting, there must be $\Pr[c' \succ_s^{w_s} c] = 1 - \Pr[c \succeq_s^{w_s} c'] \leq \frac{1}{2}$. This shows that the method is IC-R. □

However, when $|F| \geq 3$, it can no longer be IC-R because the transitivity characterized in Lemma 6.5 might be violated (see Example 7.5 for details). Finally, for the method of HEUF, it can also be IC-R for a certain class of probability distributions.

Theorem 6.7. *The GDA with the HEUF method is IC-R if and only if for any student $s \in N$, $\Pr[w_s^{f_1} \geq \mathbb{E}_{w_s \sim \mu_s}[w_s^{f_1}]] = \Pr[w_s^{f_1} \leq \mathbb{E}_{w_s \sim \mu_s}[w_s^{f_1}]]$, when the number of features $|F| = 2$.*

7 ADDITIONAL ISSUES

7.1 Restrictions of "mean=median"

Theorem 6.7 highlights a special property of those probability distributions on the first feature whose expectation equals their median when $|F| = 2$. Such distributions naturally arise in many situations. For instance, a student may firmly believe that, for her, the first feature is about twice as important as the second one, yet it is indistinguishable whether being slightly more or slightly less than twice is preferable. Moreover, we will show that these distributions possess additional theoretical properties, as they can be reduced to simple uniform distributions from certain perspectives.

Definition 7.1. We say two students s and s' are *equivalent in welfare and pairwise preferences* if they satisfy all the following:

- for any college $c \in M$, $\mathbb{E}[u_s(c)] = \mathbb{E}[u_{s'}(c)]$;
- for any feature $f \in F$, and any two colleges c_i, $c_j \in M$, $u_s^f(c_i) \geq u_s^f(c_j)$ if and only if $u_{s'}^f(c_i) \geq u_{s'}^f(c_j)$;
- for any two colleges c_i, $c_j \in M$, $\Pr[c_i \succeq_s^{w_s} c_j] \geq \Pr[c_j \succeq_s^{w_s} c_i]$ if and only if $\Pr[c_i \succeq_{s'}^{w_{s'}} c_j] \geq \Pr[c_j \succeq_{s'}^{w_{s'}} c_i]$.

Theorem 7.2. *When $|F| = 2$, then for a student s with a continuous probability μ_s that satisfies $\Pr[w_s^{f_1} \le \mathbb{E}[w_s^{f_1}]] = 1/2$, the following s' is equivalent to s in welfare and pairwise preferences:*

- *$\mu_{s'}$ is a uniform distribution over all possible weight vectors;*
- *let $A = \int_0^1 \Pr[w_s^{f_1} \le w]\mathrm{d}w$, and then for any college $c \in M$, $u_{s'}^{f_1} = 2(1-A)u_s^{f_1}$ and $u_{s'}^{f_2} = 2Au_s^{f_2}$.*

With this equivalence, the properties of a matching, such as social welfare, potential blocking pairs, and incentives for misreporting, remain unchanged. Moreover, the outcomes of methods based on these comparisons, such as HEUF and LOICV, are also preserved.

7.2 Computational Issues for More Features

In Proposition 4.2, we only give a polynomial algorithm to compute ProS(π) for a certain matching π when $|F| = 2$. Here, we supplement the computational issues when $|F| \ge 3$.

Lemma 7.3. *For any student $s \in N$ and any two colleges $c_1, c_2 \in M$, denote $\delta_s^f(c_i, c_j) = u_s^f(c_i) - u_s^f(c_j)$ for any $f \in F$, a vector of dimension $|F| - 1$ as $\vec{d}_s(c_i, c_j) = (\delta_s^{f_{|F|}}(c_i, c_j) - \delta_s^{f_k}(c_i, c_j))_{1 \le k < |F|}$, and a weight vector of the same dimension $\vec{w}_s = (w_s^f)_{1 \le f < |F|}$. We have $\Pr[c_i \succ_s^{w_s} c_j] = \Pr[\vec{d}_s(c_i, c_j)^\top \vec{w}_s < \delta_s^{f_{|F|}}(c_i, c_j)]$.*

Notice that the formula in Lemma 7.3 actually computes the cumulative probability over an $(|F|-1)$-dimensional half-space. Even with access to an oracle of c.d.f, the time complexity of computing such a probability depends on the form of the probability distribution. In general, calculating the volume of such an integral is #P-hard with respect to the number of dimensions [18], which renders the computational intractability when $|F|$ is large. Consequently, calculating ProS, the methods LOCV, LOICV, and HERF are intractable in the most general cases.

In practice, when eliciting probability distributions, it is often more realistic to consider discrete distributions with finite support, or to query the parameters of specific classes of distributions such as the uniform or normal distribution. In the former case, the computation is polynomial in the input size of the probability distribution, since we can enumerate all points located in the intersections of the target spaces. In the latter case, with specific distribution families and a relatively small number of features, we can still maintain a reasonable time complexity [21, 25] or employ standard approximation algorithms [22]. We supplement the NP-hardness of finding the highest ProS even with additional restrictions of probability distributions and utility functions as follows.

Theorem 7.4. *It is NP-hard to find a matching with the highest* ProS *in a school choice with feature-based uncertainty even when all $\mu_s \in \mu$ is uniform distribution and for each $s \in N$, $f \in F$, $c_i, c_j \in M$ such that $u_s^f(c_i) \ne u_s^f(c_j)$, it has $|u_s^f(c_i) - u_s^f(c_j)| \ge \frac{1}{m}$.*

Additionally, we have a failure of Lemma 6.5 for $|F| \ge 3$.

Example 7.5. Consider a student $s \in N$ with $u_s^{f_1}(c_1) = 0.65$, $u_s^{f_1}(c_2) = 0.91$, $u_s^{f_1}(c_3) = 0.1$; $u_s^{f_2}(c_1) = 0.9$, $u_s^{f_2}(c_2) = 0.05$, $u_s^{f_2}(c_3) = 1$; $u_s^{f_3}(c_1) = 0.21$, $u_s^{f_3}(c_2) = 0.31$, $u_s^{f_3}(c_3) = 0.7$. Then according to the formula given in Lemma 7.3, $\Pr[c_1 \succ_s^{w_s} c_2]$ equals to the probability of $(w_s^{f_1}, w_s^{f_2})$ being located right to the blue line in Figure 1, $\Pr[c_2 \succ_s^{w_s} c_3]$ equals to the probability of $(w_s^{f_1}, w_s^{f_2})$ being located above the green line, and $\Pr[c_1 \succ_s^{w_s} c_3]$ equals to the probability of $(w_s^{f_1}, w_s^{f_2})$ being located right to the yellow line. The probabilities of being located in each area are also labeled. Finally, we have $\Pr[c_1 \succ_s^{w_s} c_2] = \frac{1}{2} + \epsilon$ and $\Pr[c_2 \succ_s^{w_s} c_3] = \frac{1}{2} + \epsilon$, but $\Pr[c_1 \succ_s^{w_s} c_3] = 4\epsilon < \frac{1}{2}$.

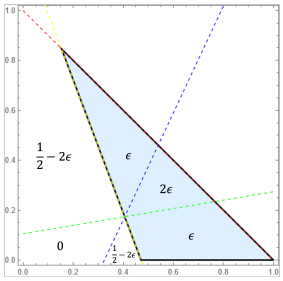

Figure 1: An example of the probability distribution of $\vec{w}_s$.

It also suggests an impossibility of IC-R algorithms when $|F| \ge 3$.

Corollary 7.6. *No matching algorithm can satisfy the property of IC-R unless it always outputs the same result when $|F| \ge 3$.*

8 DISCUSSION AND FUTURE WORK

In this paper, we study a new feature-based uncertainty model for the school choice problem. Our model assumes that a probability distribution over weight vectors is available. In practice, obtaining an arbitrary full-support distribution may be difficult, but our results do not rely on exact specifications—reasonable approximations of students' valuation uncertainty are sufficient. Such approximations are often obtainable in real systems through application data, surveys, or preference elicitation tools that evaluate programs across multiple attributes. Empirical methods such as discrete choice models and conjoint analysis can also be used to estimate preference weights from data [24]. Thus, while exact distributions may be unrealistic, structured or discretized approximations are practical and compatible with our framework. Performance may be further improved if signaling mechanisms help shape uncertainty into more structured forms.

Another realistic factor is that uncertainty may be endogenous: colleges or policymakers can influence it by signaling or information design. Although our worst-case analysis ensures robustness under such effects, it remains an open question whether improved outcomes can be achieved through explicit information design.

Finally, understanding how different uncertainty structures systematically affect matching outcomes and relate to axiomatic properties is a promising direction. Our examples suggest several patterns—for instance, more symmetric feature-weight distributions may induce acyclic comparisons, while similarity structures across features or colleges may favor different algorithms (e.g., high similarity potentially favoring LOICV). Providing a quantitative characterization of these effects is an important topic for future work.

ACKNOWLEDGMENTS

This work is supported by JST ERATO Grant Number JPMJER2301. We would also like to thank all anonymous reviewers for their efforts and insightful suggestions.

REFERENCES

[1] Atila Abdulkadiroğlu, Joshua D Angrist, Yusuke Narita, and Parag Pathak. 2022. Breaking ties: Regression discontinuity design meets market design. *Econometrica* 90, 1 (2022), 117–151.

[2] Atila Abdulkadiroğlu, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. 2018. *Impact evaluation in matching markets with general tie-breaking*. Technical Report. National Bureau of Economic Research.

[3] Atila Abdulkadiroğlu and Tayfun Sönmez. 2003. School choice: A mechanism design approach. *American economic review* 93, 3 (2003), 729–747.

[4] Akhmad Alimudin and Yoshiteru Ishida. 2021. A Study of the Random Order Mechanism for Uncertain Preferences in the Stable Marriage Problem. In *2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA)*. IEEE, 1–6.

[5] Haris Aziz, Péter Biró, Tamás Fleiner, Serge Gaspers, Ronald De Haan, Nicholas Mattei, and Baharak Rastegari. 2022. Stable matching with uncertain pairwise preferences. *Theoretical Computer Science* 909 (2022), 1–11.

[6] Haris Aziz, Péter Biró, Serge Gaspers, Ronald de Haan, Nicholas Mattei, and Baharak Rastegari. 2020. Stable matching with uncertain linear preferences. *Algorithmica* 82, 5 (2020), 1410–1433.

[7] Haris Aziz, Serge Gaspers, Zhaohong Sun, and Toby Walsh. 2019. From Matching with Diversity Constraints to Matching with Regional Quotas. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*. 377–385.

[8] Haris Aziz, Isaiah Iliffe, Bo Li, Angus Ritossa, Ankang Sun, and Mashbat Suzuki. 2024. Envy-free house allocation under uncertain preferences. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 9477–9484.

[9] Haris Aziz, Venkateswara Rao Kagita, Baharak Rastegari, and Mashbat Suzuki. 2024. Approval-based committee voting under uncertainty. *arXiv preprint arXiv:2407.19391* (2024).

[10] Angela D Bell, Heather T Rowan-Kenyon, and Laura W Perna. 2009. College knowledge of 9th and 11th grade students: Variation by school and state context. *The Journal of Higher Education* 80, 6 (2009), 663–685.

[11] Sung-Ho Cho, Taiki Todo, and Makoto Yokoo. 2022. Two-Sided Matching over Social Networks. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*. International Joint Conferences on Artificial Intelligence Organization.

[12] Yaqin Chu, Junjie Luo, and Tianyang Zheng. 2024. Stable matching with approval preferences under partial information. In *International Conference on Algorithmic Aspects in Information and Management*. Springer, 64–75.

[13] Tom Dvir, Renana Peres, and Zeév Rudnick. 2020. Modelling the expected probability of correct assignment under uncertainty. *Scientific Reports* 10, 1 (2020), 15080.

[14] David Gale and Lloyd S Shapley. 2013. College admissions and the stability of marriage. *The American Mathematical Monthly* 120, 5 (2013), 386–391.

[15] Guillaume Haeringer and Flip Klijn. 2009. Constrained school choice. *Journal of Economic theory* 144, 5 (2009), 1921–1947.

[16] Justine S Hastings, Richard Van Weelden, and Jeffrey M Weinstein. 2007. Preferences, information, and parental choice behavior in public school choice.

[17] Caroline M Hoxby and Sarah Turner. 2015. What high-achieving low-income students know about college. *American Economic Review* 105, 5 (2015), 514–517.

[18] Leonid Khachiyan. 1993. Complexity of polytope volume computation. In *New trends in discrete and computational geometry*. Springer, 91–101.

[19] Kei Kimura, Kweiguu Liu, Zhaohong Sun, Kentaro Yahiro, and Makoto Yokoo. 2025. Multi-stage generalized deferred acceptance mechanism: Strategyproof mechanism for handling general hereditary constraints. *Autonomous Agents and Multi-Agent Systems* 39, 2 (2025), 1–25.

[20] Laura Owen, Timothy A Poynton, and Raeal Moore. 2020. Student preferences for college and career information. *Journal of college access* 5, 1 (2020), 7.

[21] Franco P. Preparata and David E. Muller. 1979. Finding the intersection of n half-spaces in time O (n log n). *Theoretical Computer Science* 8, 1 (1979), 45–55.

[22] James Ridgway. 2016. Computation of Gaussian orthant probabilities in high dimension. *Statistics and computing* 26, 4 (2016), 899–916.

[23] Alvin E Roth. 2008. Deferred acceptance algorithms: History, theory, practice, and open questions. *international Journal of game Theory* 36, 3-4 (2008), 537–569.

[24] Geoffrey N Soutar and Julia P Turner. 2002. Students' preferences for university: A conjoint analysis. *International journal of educational management* 16, 1 (2002), 40–45.

[25] Hong-Jie Sun. 1988. A general reduction method for n–variate normal orthant probability. *Communications in Statistics-Theory and Methods* 17, 11 (1988), 3913–3921.

[26] Ryota Takeshima, Kei Kimura, Ayumu Kuroki, Temma Wakasugi, and Makoto Yokoo. 2025. A New Relaxation of Fairness in Two-Sided Matching Respecting Acquaintance Relationships. In *28th European Conference on Artificial Intelligence, ECAI 2025, including 14th Conference on Prestigious Applications of Intelligent Systems, PAIS 2025*. IOS Press BV, 3727–3734.