Université de Paris

École Doctorale Sciences Mathématiques de Paris Centre

Institut de Recherche en Informatique Fondamentale

Thèse de Doctorat en Informatique

Analysis of Random Models for Stable Matchings

Candidate:
Simon MAURAS

Advisors:
Claire MATHIEU
Hugo GIMBERT

A thesis submitted in fulfillment of the requirements for the Degree of Doctor of Philosophy (Ph.D.)

And considered on December 15th, 2021, by a jury comprised of:

Claire Mathieu	Research Director at CNRS, IRIF, Paris, France	*Advisor*
Hugo Gimbert	Researcher at CNRS, LaBRI, Bordeaux	*Co-Advisor*
Amos Fiat	Professor at TAU, Tel Aviv, Israel	*Referee*
Olivier Tercieux	Professor at PSE, Paris, France	*Referee*
Michal Feldman	Professor at TAU, Tel Aviv, Israel	*Examiner*
Nicole Immorlica	Senior Researcher at MSR, New England, USA	*Examiner*
Stéphan Thomassé	Professor at ENS de Lyon, France	*Examiner*
Tristan Tomala	Professor at HEC, Paris, France	*Examiner*

Acknowledgments

Je tiens tout d'abord à remercier Claire Mathieu et Hugo Gimbert, qui m'ont accompagné tout au long de ces 3.33 années de thèse. Je me considère extrêmement chanceux d'avoir pu découvrir le monde de la recherche en votre compagnie. Merci pour le temps que vous m'avez consacré, entre séances à Paris, visites à Bordeaux et réunions en ligne. J'espère que vous avez apprécié travailler avec moi autant que j'ai apprécié travailler avec vous!

I would also like to express my gratitude to Amos Fiat an Olivier Tercieux who took the time to read and carefully review this thesis; and to Michal Feldman, Nicole Immorlica, Stéphan Thommassé and Tristan Tomala who accepted to be part of the committee.

During my PhD, I had the unique opportunity to attend scientific programs and seminars to help me grasp the many interesting connections between Computation (from where I was coming) and Economics (which was completely new to me). For this reason, I would like to thank Jacob Leshno who invited me to the "Matching-Based Market Design" program at Simons institute in Berkeley; Bettina Klaus, Péter Biro and Tamás Fleiner who organized the "Matching under Preferences" seminar in Dagstuhl; Vianney Perchet who invited me to a "Machine Learning and Economics" session at CMStatistics; and Olivier Tercieux and Julien Combes who organize a "Matching Reading Group" in Paris; together with all the amazing people that I had the chance to meet at those aforementioned programs.

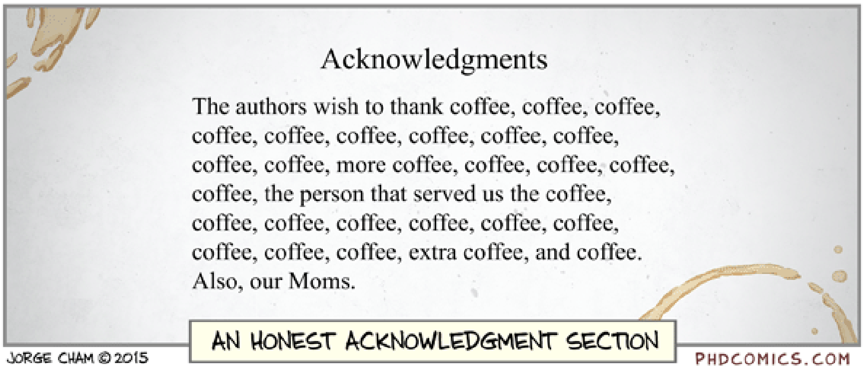
Although not included in this manuscript, I spent a portion of my PhD working on exciting side projects. I would like to thank my (wonderful) past, current and future collaborators: Vincent Cohen-Addad, Guillaume Duboc, Max Dupré la Tour, Paolo Frasca, Lulla Opatowski and Laurent Viennot; Chien-Chung Huang, Naonori Kakimura and Yuichi Yoshida; Pierre Aboulker, Spyros Angelopoulos, Evripidis Bampis, Thomas Bellitto, Bruno Escoffier and Michel de Rougemont; Michal Feldman, Amos Fiat, Federico Fusco, Stefano Leonardi and Rebecca Reiffenhäuser.

This PhD is one milestone in my research journey, which would not have started without all the teachers that I had the chance to listen to during my studies in Nantes, Lyon and Paris (unfortunately, this thesis is to short for me to give an exhaustive list). In addition to them, I must thank Mathias Hiron who introduced me to the world of algorithms, and Michael Bender who supervised me during a master's internship and from whom I learned most of the tools for the randomized analysis of algorithms that are used in this thesis.

I am deeply grateful to my friends and colleagues who contributed to this thesis through many interesting discussions, and without whom my PhD would have been much less fun :-) I am grateful to the administrative team at IRIF (Dieneba, Etienne, Eva, Ines, Jemuel, Natalia, Omur, Sandrine) for their help and their cheerfulness; to my officemates (Gabriel, Klara, Olivier, Pierre, Yassine, Yixin, Zhouningxin) to whom I imposed too many coffee breaks; to my lunchmates (Amaury, Abhicheck, Anupa, Avinandan, Daniel, Mikael, Patrick, Pierre, Robin, Sander, Simona) for our daily trips to the ministry of sports; to my conferencemates (David and Mathieu) for you friendship; and more genarally to all of my mathematician, computer scientist and (even) physicist friends who I met in Lyon, Paris or Bordeaux.

Je voudrais finir en remerciant tous mes proches. Merci Louisiane de partager ta vie avec moi (tu as également les remerciements de mon jury de thèse, qui sans toi auraient du lire un manuscrit beaucoup trop long). Merci Apolline, Capucine, Martin et Simba pour les bons moments en votre compagnie (et pour les heures que nous avons passés sur un calcul de déterminant). Enfin, merci Papa, Céline, Cyrielle, Enora, Lauriane, Pauline, Thomas et Zoé, merci Maman, et merci à toute ma famille pour votre bonne humeur infaillible (sauf parfois quand je pose une énigme mathématique à table).

Finally, to everyone who is reading this and I might have forgotten, thank you!



Résumé

Dans un marché biparti, deux types d'agents ont des préférences sur les agents du côté opposé. Parmi les exemples classiques on retrouve l'affectation d'étudiants dans des universités, de docteurs dans des hôpitaux, de travailleurs à des offres d'emploi et, dans l'analogie historique des mariages stables, l'appariement d'hommes et de femmes. Dans un article fondateur, Gale et Shapley introduisent la procédure d'acceptation différée, dans laquelle un côté propose et l'autre côté dispose, permettant de calculer un matching stable.

Les matchings stables constituent un sujet de recherche important en informatique et en économie. Des résultats issus de littérature informatique décrivent la structure de treillis complet de l'ensemble des matchings stables, ainsi que les algorithmes permettant de le calculer. Dans la littérature économique ont été étudiées les questions de manipulabilité par les agents participant à un marché biparti, à la fois du point de vue théorique et empirique.

Une série récente de travaux étudie les propriétés des matchings stables, en utilisant des modèles stochastiques dans lesquels les préférences des agents sont générées aléatoirement. Cette thèse poursuit cette approche, et considère deux questions : "qui peut manipuler ?" et "qui obtient quoi ?".

La première partie, abordant la question "qui peut manipuler", contient trois résultats différents. Dans un premier résultat (Chapter 4), nous montrons que lorsque les agents d'un des côtés du marché ont des préférences très corrélées, les opportunités de manipulabilité sont réduites. Dans un second résultat (Chapter 5), nous montrons que des préférences décorrélées constituent un pire cas. Les preuves de ces deux résultats sont basées sur une analyse probabiliste de l'algorithme calculant les opportunités de manipulabilité. Dans un troisième résultat (Chapter 6), nous étudions le jeu à information incomplète où des étudiants peuvent postuler à un nombre limité d'écoles et, par conséquent, choisissent leur liste de préférence de manière stratégique. Nous prouvons l'existence d'un équilibre symétrique et proposons des algorithmes permettant de le calculer dans plusieurs cas particuliers.

La seconde partie, abordant la question "qui obtient quoi ?", contient également trois résultats. Dans un premier résultat (Chapter 7), nous montrons que sous certaines conditions sur la distribution d'entrée sur les préférences, les deux variantes de l'algorithme d'acceptation différée produisent exactement la même distribution de sortie sur les matchings. Les preuves utilisent la structure de treillis de l'ensemble des matchings stables, montrent qu'un matching fixé a la même probabilité d'être la borne inférieure ou supérieure, et donnent une formule pour la probabilité que deux agents soient appariés. Dans un second résultat (Chapter 8), nous considérons un modèle dans lequel la probabilité que deux agents s'apprécient est quantifiée par une matrice de "popularités", et nous expliquons que les probabilités d'appariement sont asymptotiquement données par la matrice renormalisée dont les lignes/colonnes ont une somme égale à 1. Dans un troisième résultat (Chapter 9), nous étudions la complexité de l'algorithme d'acceptation différée, qui se rapporte à l'étude du rang que chaque agent donne à son partenaire. Les preuves sont basées sur une réduction au problème de collection de coupons.

Summary

In a two sided matching market, two types of agents have preferences over one another. Examples include college admissions (students and colleges), residency programs (doctors and hospitals), job markets (workers and jobs) and, in the classical analogy, stable marriages (men and women). In a founding paper, Gale and Shapley introduced the deferred acceptance procedure, where one side proposes and the other disposes, which computes a stable matching.

Stable matchings have been an extensive research topic in computer science and economics. Results in the computer science literature include the lattice structure of the set of stable matchings, and algorithms to compute it. In the economics literature, researcher have studied the incentives of agents taking part in two-sided matching markets, both from the theoretical and empirical point of views.

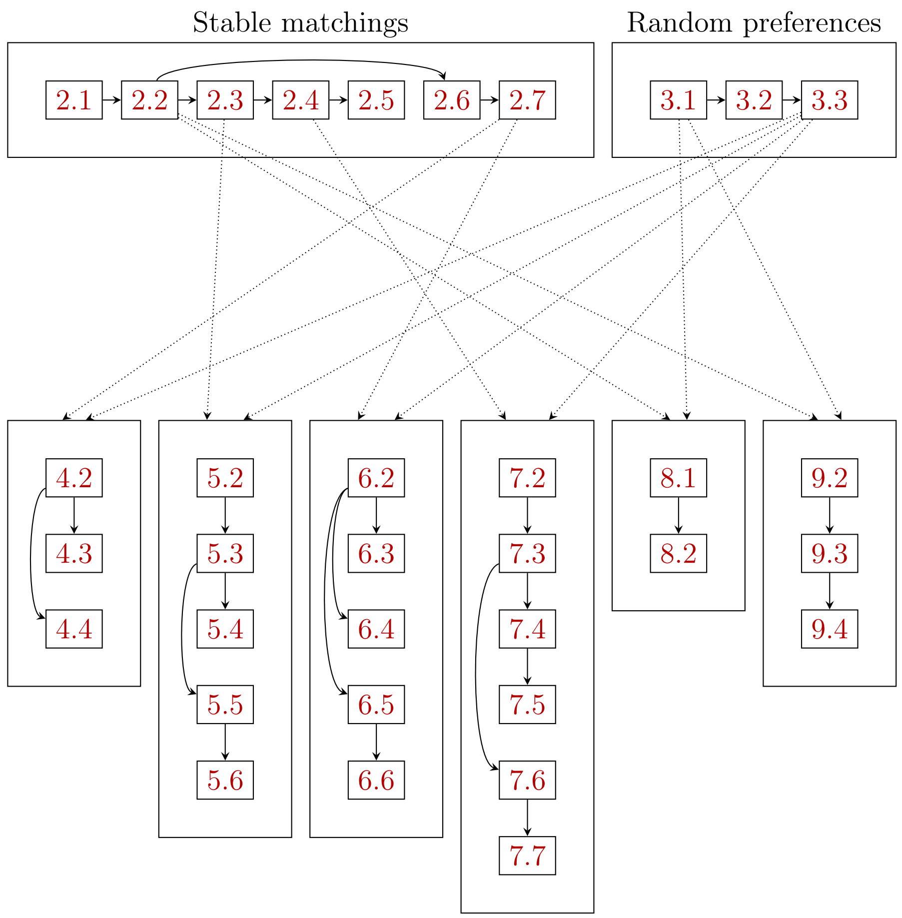
A recent line of works study the properties of stable matchings, using stochastic models of two-sided matching markets where the preferences of agents are drawn at random. This thesis follows this direction of inquiry, and considers two main questions: "who can manipulate?" and "who gets what?".

The first part, addressing the question "who can manipulate?", contains three different results. In a first result (Chapter 4), we show that when one side of the market has strongly correlated preferences, incentives to manipulate are reduced. In a second result (Chapter 5), we show that uncorrelated preferences is a worst case situation when compared to correlated preferences. Proofs of both results are based on a randomized analysis of the algorithm which computes the incentives agents have to manipulate. In a third result (Chapter 6), we study the incomplete information game where students must apply to a limited number of schools, and thus report their preferences strategically. We prove the existence of symmetric equilibria and design algorithms to compute equilibria in various special cases.

The second part, addressing the question "who gets what?", also contains three different results. In a first result (Chapter 7), we show that under a certain input distribution of preferences, the two variants of deferred acceptance produce the same output distribution on matchings. Proofs use the lattice structure of stable matchings, show that a fixed matching has the same probability of being the top or bottom element, and give a closed formula for the probability of two agents being matched. In a second result (Chapter 8), we consider a model where the probabilities that agents like each are quantified by a "popularity" matrix, and we give evidences that the probabilities that deferred acceptance matches agents is asymptotically given by the scaled matrix where lines/columns sum up to 1. In a third result (Chapter 9), we study the time complexity of deferred acceptance, which relates to the rank people from the proposing side give to their partner. Proofs are based on a reduction to the coupon collector's problem.

Contents

Dependency Diagram



List of Notations

	Set theory
$[n]$	set of integers $\{1, 2, \ldots, n\}$
$f^{-1}(Y)$	preimage $X = \{x \in A \mid f(x) \in Y\}$ of Y under the function $f : A \to B$
$\mathbb{1}[\text{predicate}]$	indicator function, equal to 1 if the predicate is true, and 0 otherwise
	Probability and measure theory
$(\Omega, \Sigma, \mathbb{P})$	underlying probability space
$B \in \mathcal{B}(S)$	Borel subset of the topological space S
$\mu \in \Delta(S)$	Borel probability measure over the topological space S
$X \sim \mu$	random variable $X : \Omega \to S$ with distribution μ
$\mathbb{P}[X \in B]$	probability that X is in B, equal to $\mathbb{P}(X^{-1}(B)) = \mu(B)$
$\mathbb{E}[X]$	expected value of X, equal to $\int_x x \cdot \mathrm{d}\mu(x)$
$\mathrm{Unif}(S) \in \Delta(S)$	uniform distribution over S, when it exists
$\mathrm{Geom}(p) \in \Delta(\mathbb{N}^*)$	geometric distribution with parameter $p \in (0, 1)$
$\mathrm{Binom}(n, p) \in \Delta(\{0, 1, \ldots, n\})$	binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0, 1]$
$\mathrm{Exp}(\lambda) \in \Delta(\mathbb{R}_+)$	exponential distribution with parameter $\lambda \in \mathbb{R}_+^*$

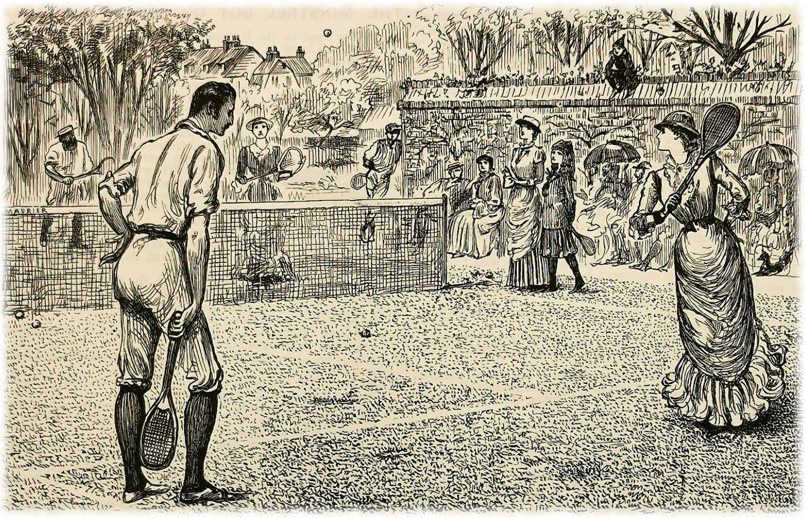
Part I

Preliminaries

1 | Introduction

1.1 The tale of a tennis tournament

Imagine that you are organizing a mixed double tennis tournament: M men and W women have registered, and will soon report preferences over partners with whom they want to play[1]. You are in charge of forming pairs: each team must be comprised of one man and one woman, who both agree to play with each other. From experience, you know that when checking the composition of teams, a man and a woman will complain if they are not matched together and prefer each other to their respective partners (such matching would be unstable).



Amenities Of The Tennis Lawn, George du Maurier, 1883

Fortunately, being a seasoned tournament organizer, you are well accustomed to the theory of stable matchings. Even better, having already organized similar tournaments in the past, you have some prior (distributional) knowledge on the preferences of every player! Informally speaking, the preferences of contestants are correlated: good players are very popular (one-sided correlations), and players with similar styles perform well together (cross-sided correlations).

The doors of the tennis stadium are now opening, and every contestant arrives with multiple petty questions: "Should I report my preferences truthfully?", "With whom am I most likely to be matched?". In this thesis, we try to answer those questions.

[1]In the original analogy, each person ranks those of the opposite sex in accordance with their preferences for a marriage partner. This "sportive" rephrasing of the classical model is due to Ágnes Cseh and Jannik Peters.

1.2 Random two-sided matching markets

Two-sided matching markets describe matching scenarios between to types of agents, where agents from both sides of the market have ordered preferences over the opposite side. Examples include college admissions (students and colleges), residency programs (doctors and hospitals), job markets (workers and jobs), and of course, mixed double tennis tournaments (men and women). In each setting, a centralized clearinghouse computes an allocation, pairing agents from opposite sides of the market. Instability occurs when two agents who are not matched together prefer each other to the partners they have been paired with by the clearinghouse. Such instability might be a cause of market failure [Rot08], where agents leave the market to arrange better matches. In a founding paper, Gale and Shapley [GS62] designed the deferred acceptance mechanism, where one side propose while the other dispose, which they prove always outputs a stable matching. Since then, many matching markets have successfully implemented mechanisms based on a deferred acceptance procedure [RP99; APR05; Abd+05; Cor+19].

Stable matchings have been an extensive research topic in computer science and economics. Results in the computer science literature include the lattice structure of the set of stable matchings, and algorithms to compute it. In the economics literature, researcher have studied the incentives of agents taking part in two-sided matching markets, both from the theoretical and empirical point of views. The literature of matching under random preferences came from the need to model real matching markets, and give formal proofs of properties observed empirically.

1.3 Contributions of this thesis

In the preliminary part of this thesis, Chapter 2 surveys classical results about stable matchings, and Chapter 3 discuss random models of preferences. Each subsequent chapter corresponds to a research question. Whenever possible, chapters contains an introduction with the main results, related works and one takeaway observation; a section with computer simulations illustrating the main theorems; and a conclusion with open questions and future directions.

Most papers from the literature of matching under random preferences study either the incentive compatibility or the outcomes of matching procedures. Parts II and III of this thesis consider the questions “who can manipulate?” and “who gets what?”.

1.3.1 Who can manipulate?

We survey incentive compatibility results of stable matching procedures in Section 2.7. In particular, if an agent lies about their preference list, this gives rise to new stable matchings, where they will be no better off than they would be in one of the original stable matching. Thus, a person can only gain from strategic manipulation up to the maximum difference between their best and worst partners in stable matchings, and an agent who has a unique stable partner does not have incentives to misreport their true preference list.

Core-convergence. In a renowned paper, Roth and Peranson [RP99] report that in 5 years of data from the US National Resident Matching Program (which matches approximately 20000 doctors to hospitals each year), exactly 4 doctors had multiple stable partners, and thus could have benefited from misreporting their preferences. This phenomenon is often called core-convergence, in reference to the core of a cooperative game, which here corresponds to the set of stable matchings. Theoretical explanations for core-convergence have been given by Immorlica and Mahdian [IM15], and Ashlagi, Kanoria and Leshno [AKL17], based on the fact that agents report constant size preference lists, and that markets are typically slightly unbalanced. In Chapter 4, we explore an alternative explanation, based on the fact that agents have correlated preferences.

Weak core-convergence. As a point of comparison, Knuth, Motwani and Pittel [KMP90; Pit92] study balanced markets where agents have complete preferences drawn uniformly at random, and show that an agent has $\sim \ln N$ stable partners. This result does not imply strategy-proofness (achieved with 1 stable partner), but is much smaller than the worst case situation where agents have N stable partners. Thus, we will refer to having few stable partners as weak core-convergence. Chapter 5 considers matching markets with non-uniform distributions, and shows that positive correlations between the preferences of agents helps weak core-convergence.

Choice with constraints. Back to the US National Resident Matching Program, Echenique, Gonzalez, Wilson and Yariv [Ech+20] make the puzzling observation that a large majority of doctors are matched with one of their top (reported) choices, whereas surveys indicate that they have similar preferences over hospitals. They explain this observation by the fact that before reporting their preferences, doctors and hospitals interact in a decentralized interview process. Because the number of interview is limited, they argue that reported preferences should not be taken at face value. In Chapter 6, we study an incomplete information game where applicants must behave strategically because of an upper quota on the number of applications they can submit.

1.3.2 Who gets what?

In a two-sided matching market, multiple stable matchings can exist. The deferred acceptance produces a matching which is optimal for the proposing side, and pessimal for the receiving side. More importantly, the deferred acceptance mechanism is not Pareto efficient, in the sense that agents from the proposing side might all weakly prefer another (unstable) matching. This motivates the use of more efficient yet possibly unstable mechanisms (top-trading cycles or serial dictatorship), for example when allocating students to public schools [Abd+20]. In a founding paper of the school choice literature, Abdulkadiroglu and Sönmez [AS03] discuss mechanism design as a trade-off between the quality of outcomes and other desirable properties.

Ex-post outcomes. When considering random models of two-sided matching markets, the rank or utility each person gives to their partner corresponds to *ex-post* outcomes, that is after having drawn each person's preferences. Wilson, Knuth and Pittel [Wil72; Knu76; Pit89] show that in balanced markets with uniformly random preferences, the deferred acceptance procedure match proposing agents to one of their top $\sim \ln N$ choice, and receiving agents to one of their $\sim N/\ln N$ choices, in expectation. Interestingly, Knuth [Knu96] shows that $\sim \ln N$ is also equivalent to the expected rank obtained by the Pareto efficient procedures discussed above.

Conversely, when agents from each side have identical preferences, the outcome is assortative, in the sense that each person is matched with someone of corresponding rank. Lee [Lee16] discusses a model where preferences are induced by vertical cardinal utilities, and Chapter 4 explores the situation where agents have strongly correlated (yet non-identical) ordinal preferences.

Ex-ante outcomes. When agents have heterogeneous random preferences, one can take a different approach and look at the probability that two agents will be matched. This corresponds to *ex-ante* outcomes, that is from the point of view of a mechanism designer who only has prior knowledge (for example from historical data) on the preferences of agents. Chapters 7 and 8 give exact and asymptotic formula for the match probabilities under certain preference distributions.

Speed of deferred acceptance. Finally, it is important to observe that the time complexity of the deferred acceptance procedure is equal to the sum of ranks agents from the proposing side give to their eventual partners. This was in fact the original motivation of Wilson and Knuth [Wil72; Knu76], whose result (discussed above) imply that the average complexity of deferred acceptance is $\sim N \ln N$. However, such notion of "sequential" complexity is not adapted to recent implementations of the deferred acceptance mechanism. In the new college admission system in France, students interact with the platform, choosing which offer they want to keep when receiving proposals from multiple schools. In an alternative description of the deferred acceptance algorithm, each agent from the proposing side can send of offer per day. Chapter 9 looks at the expected number of days required by the procedure when agents have uniform preferences.

1.4 Proof techniques.

Stochastic models provide interesting proof techniques when coupled with structural and algorithmic results on the set of stable matchings. Technical contributions of this thesis have been split in chapters based on the following notions.

Stability. The definition of stability itself (see Section 2.1) already provides interesting insights. Knuth, Pittel and co-authors [Knu76; Pit89; Pit92; PSV07; LP09; Pit18] write the probability of stability of a matching as an integral formula, and use it to compute various quantities of interest (number of stable matchings, number and rank of stable partners). Integral formulae are used in [LY14] to measure the efficiency of stable matchings. Proofs in Chapter 7 use integral formulae to compute the output distribution of deferred acceptance.

Deferred acceptance. The deferred acceptance procedure (see Section 2.2) can be analyzed as a stochastic process where agents draw their preferences online. Wilson and Knuth [Wil72; Knu76] compute the average complexity of deferred acceptance via a reduction to the coupon collector's problem. It is used in [KMQ21; Ash+21] to compute the average rank of each person's partner as a function of the parameters of the model. Chapters 8 and 9 are based on such techniques.

Stable partners. Enumerating stable partners (see Section 2.3) can be done via an extension of the deferred acceptance mechanism. Knuth, Motwani and Pittel [KMP90] gave a stochastic analysis of this algorithm, which was later extended in [IM15; AKL17] to show that almost everyone has a unique stable partner, in random markets where some agents remain single. Proofs in Chapters 4 and 5 are based on such techniques.

Lattice structure. Finally, the set of stable matchings has a lattice structure (see Section 2.4), that can be used to derive additional properties. Pittel's integral formulae [Pit92] give asymptotic value for the number and size of rotations (symmetric difference of two consecutive stable matchings). A recent work [NNV21] makes a connection between the statistics of random mappings and of rotations that are exposed in a matching. Proofs from Chapter 7 partially rely on the lattice structure of stable matchings.

2 | Stable Matchings

The stable matching problem was first defined by Gale and Shapley [GS62] and later extended to allow sets of agents of different sizes [MW70] and incomplete preferences [GI89]. Results presented in this chapter can be found in the books of Knuth [Knu76; Knu97], Gusfield and Irving [GI89], Roth and Sotomayor [RS92], Manlove [Dav13], and Echenique, Immorlica and Vazirani [EIV22]. One novelty in this chapter is the presentation stable matchings as a sub-lattice of the lattice of stable permutations (which have been defined for the more general stable roommates problem).

2.1 Formal definitions

Let us start with formal definitions and classical notations. Let $\mathcal{M} = \{m_1, \ldots, m_M\}$ be a set of M men, $\mathcal{W} = \{w_1, \ldots, w_W\}$ be a set of W women, and $N = \min(M, W)$. Each person declares which members of the opposite sex they find acceptable, then gives a strictly ordered preference list of those members. Preference lists are *complete* when no one is declared unacceptable.

Definition 2.1 (Preference profile)**.** Formally, we represent the preference list of a man m as a total order $\succ_m$ over $\mathcal{W} \cup \{m\}$, where $w \succ_m m$ means that man m finds woman w acceptable, and $w \succ_m w'$ means that man m prefers woman w to woman w'. Similarly we define the preference list $\succ_w$ of woman w. A preference profile is a tuple containing the preference list of each person.

Our goal is to match men and women, in such a way that no man-woman pair prefer each other to their respective partners, which would create instability.

Definition 2.2 (Stable matching)**.** Formally, a matching is a function $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W}$, which is self-inverse ($\mu^2 = \mathrm{Id}$), where each man m is paired either with a woman or himself ($\mu(m) \in \mathcal{W} \cup \{m\}$), and symmetrically, each woman w is paired with a man or herself ($\mu(w) \in \mathcal{M} \cup \{w\}$).

A man-woman pair (m, w) is blocking matching μ if $m \succ_w \mu(w)$ and $w \succ_m \mu(m)$. Abusing notations, observe that μ matches a person p with an unacceptable partner when p would prefer to remain single, that is when the pair (p, p) is blocking. A matching with no blocking pair is stable.

We will use the terms of stable pair for a pair which belongs to at least one stable matching. Correspondingly, we call stable partner (resp. stable wife, stable husband, ...) any partner (resp. partner, wife, husband, ...) with whom someone is matched in a stable matching.

2.2 Deferred acceptance algorithm

In their founding paper, Gale and Shapley [GS62] introduce the men-proposing deferred acceptance mechanism, which they show computes a men-optimal stable matching. In subsequent works, McVitie and Wilson [MW70], and Gusfield and Irving [GI89] extends those results to unbalanced markets with incomplete preferences.

Algorithm 2.1 Men Proposing Deferred Acceptance.

Input: Preferences of men $(\succ_m)_{m \in M}$ and of women $(\succ_w)_{w \in \mathcal{W}}$.
Initialization : Start with an empty matching μ.
While a man m is single and has not proposed to every woman he finds acceptable**, do**
  m proposes to his favorite woman w he has not proposed to yet.
  If m is w's favorite acceptable man among all proposals she received**, then**
    w accepts m's proposal, and rejects her previous husband if she was married.
Output: Resulting matching.

Theorem 2.3 (Adapted from [GS62])**.** *Algorithm 2.1 outputs a stable matching $\mu_{\mathcal{M}}$ independently of the order in which men are chosen to propose. In the matching $\mu_{\mathcal{M}}$, every man (resp. woman) has his best (resp. her worst) stable partner. Symmetrically, there exists a stable matching $\mu_{\mathcal{W}}$ in which every woman (resp. man) has her best (resp. his worst) stable partner.*

$$\begin{aligned}\forall \mu \text{ stable matching}, \quad &\forall m \in \mathcal{M}, \quad \mu_{\mathcal{M}}(m) \succeq_m \mu(m) \succeq_m \mu_{\mathcal{W}}(m)\\ &\forall w \in \mathcal{W}, \quad \mu_{\mathcal{W}}(w) \succeq_w \mu(w) \succeq_w \mu_{\mathcal{M}}(w)\end{aligned}$$

Proof. First, we show that the output $\mu_{\mathcal{M}}$ of Algorithm 2.1 is stable. Men never propose to women they find unacceptable, and women never accept proposals from men they find unacceptable, thus each person is either single or matched with an acceptable partner. For every pair (m, w), man m proposed to every women he prefers to $\mu_{\mathcal{M}}(m)$, and woman w prefer $\mu_{\mathcal{M}}(w)$ to every other proposal she received, thus (m, w) cannot be a blocking pair.

Second, we show that every man receives his favourite stable partner. Let μ be a stable matching. If for some man m we have $\mu(m) \succ_m \mu_{\mathcal{M}}(m)$, then $\mu(m) \neq m$. For the sake of contradiction, let m be the first man (such that $\mu(m) \neq m$) who is rejected by $w = \mu(m)$ during the execution of Algorithm 2.1, and let m' be the best proposer to w at that time. By construction, m' has not yet been rejected by $\mu(m')$, thus he prefers w to $\mu'(m)$. Because $m' \succ_w \mu(w)$ and $w \succ_{m'} \mu(m')$, the pair (m', w) blocks matching μ, which is a contradiction.

Third, we show that every woman receives her least favourite stable partner. Let μ be a stable matching. For the sake of contradiction, assume that $\mu_{\mathcal{M}}(w) \succ_w \mu(w)$ for some woman w. Then $\mu_{\mathcal{M}}(w) \neq w$, and we define $m = \mu_{\mathcal{M}}(w)$. By optimality of $\mu_{\mathcal{M}}$, man m prefers w to $\mu(m)$. Thus, the pair (m, w) blocks matching μ, which is a contradiction.

Fourth, because Algorithm 2.1 always outputs a matching where each man receives his favourite stable partners, then the output is independent of the order in which men are chosen to propose. ☐

Theorem 2.4 (Adapted from [MW70])**.** *Each person is either matched in all stable matchings, or single in all stable matchings. In particular, a woman is matched in all stable matchings if and only if she received at least one acceptable proposal during Algorithm 2.1.*

Proof. Denote $\mu(\mathcal{M}) \cap \mathcal{W}$ (resp. $\mu(\mathcal{W}) \cap \mathcal{M}$) the set of women (resp. men) who are matched in a matching μ, and observe that $|\mu(\mathcal{M}) \cap \mathcal{W}| = |\mu(\mathcal{W}) \cap \mathcal{M}|$ by a pigeonhole principle. For every stable matching μ, Theorem 2.3 shows that

$$\mu_{\mathcal{M}}(\mathcal{W}) \cap \mathcal{M} \supseteq \mu(\mathcal{W}) \cap \mathcal{M} \supseteq \mu_{\mathcal{W}}(\mathcal{W}) \cap \mathcal{M} \tag{2.1}$$

$$\mu_{\mathcal{W}}(\mathcal{M}) \cap \mathcal{W} \supseteq \mu(\mathcal{M}) \cap \mathcal{W} \supseteq \mu_{\mathcal{M}}(\mathcal{M}) \cap \mathcal{W} \tag{2.2}$$

Equation (2.1) shows that $\mu_{\mathcal{M}}$ matches at least as many persons as $\mu_{\mathcal{W}}$, and conversely Equation (2.2) shows that $\mu_{\mathcal{W}}$ matches at least as many persons as $\mu_{\mathcal{M}}$. By a cardinality argument, every inclusion is an equality, which concludes the proof. □

2.3 Algorithm for enumerating stable partners

In subsequent works, McVitie [MW71] and Gusfield [Gus87] extended the deferred acceptance procedure to enumerate stable matchings and stable partners: by breaking existing marriages and waiting for the algorithm to converge, one can reach every stable matching. We present here a simplified version, similar to the one of Knuth, Motwanni and Pittel [KMP90].

Starting from the men-optimal stable matching (Algorithm 2.1), Algorithm 2.2 continues the execution of the deferred acceptance procedure, in which w rejects every proposal she receives. Stable husbands of w are "best so far" proposals, that is men who proposed to w and were preferred to all men who proposed before them. The proof relies on Lemma 2.6 below.

Algorithm 2.2 Extended Men Proposing Deferred Acceptance.

Input: Preferences of men $(\succ_m)_{m \in M}$ and of women $(\succ_w)_{w \in \mathcal{W}}$. Fixed woman $w^* \in \mathcal{W}$.
Initialization: Start by executing Algorithm 2.1, if w^* is unmatched then stop.
Phase 1: sequence of proposals
Let $m \leftarrow \mu_{\mathcal{M}}(w^*)$ be the *proposer*, let $S \leftarrow [\mu_{\mathcal{M}}(w^*)]$ be the **sequence of proposals** to w^*.
While the *proposer* m has not proposed to every woman he finds acceptable**, do**
  m proposes to his favorite woman w he has not proposed to yet.
  If $w = w^*$**, then**
    append m to the sequence S
  else if w has never been matched **then**
    break the while loop,
  else if m is w's favorite acceptable man among all proposals she received **then**
    w rejects her previous husband m', accepts m, the *proposer* becomes m'
Phase 2: stable husbands
For each proposal $m \in S$ made to w^*, in order of reception**, do**
  If m is the best proposition w received so far**, then**
    m is a stable husband of w^*.
Output: Set of stable husbands of w^*.

Theorem 2.5 (Adapted from [KMP90])**.** *Algorithm 2.2 outputs w^*'s set of stable husbands.*

Proof. First, we show that every man m returned by Algorithm 2.2 is a stable partner of woman w^*. Let μ_m be the matching when w^* received the proposal from m during the execution of Algorithm 2.2, in particular $\mu_m(w^*) = m$. Then, μ_m is exactly the matching returned by Algorithm 2.1 if w^* truncates her preference list and declares unacceptable every man to whom she prefers m, which proves that μ_m is stable under the truncated preferences of w^*. Lemma 2.6 proves that μ_m is stable.

Second, we show that every stable partner of w^* has been found by Algorithm 2.2. Let μ be a stable matching, and let $m = \mu(w^*)$. Let μ_m be the matching returned by Algorithm 2.1 if w^*

truncates her preference list and declares unacceptable men to whom she prefers m. Lemma 2.6 proves that μ is stable when w^* truncates her preferences. Because μ_m is women-pessimal, we have $m = \mu(w^*) \succeq \mu_m(m) \succeq m$, and thus $\mu_m(w^*) = m$. We now run Algorithm 2.1 with truncated preferences, but we delay $\mu_{\mathcal{M}}(w^*)$ proposing to w^* as much as possible: the first part (before $\mu_{\mathcal{M}}(w^*)$ proposes to w^*) exactly corresponds to Algorithm 2.1 with original preferences; and the second part (after $\mu_{\mathcal{M}}(w^*)$ proposed to w^*) exactly corresponds to Algorithm 2.2 with original preferences, up to the point when m proposes to w^* (who accepts). Thus, m will propose to w^* in Algorithm 2.2 with original preferences, at which point it is the best proposition received, which concludes the proof. □

Lemma 2.6. *Let w^* be a fixed women and let m^* be a fixed man. Let μ be a matching such that $\mu(w^*) \succeq_{w^*} m^*$. Then μ is stable if and only if it is stable when w^* truncates her preferences and declares unacceptable men to whom she prefers m^*.*

Proof. For every m, observe that $m \succ_{w^*} \mu(w^*)$ is independent from whether or not preferences are truncated after m^*, thus a pair is blocking with the truncated preferences if and only if it is blocking with the original preferences. □

2.4 Lattice of stable matchings

The set of stable matchings can be partially ordered using the preferences of agents:

$$\begin{aligned} \forall \mu_1, \mu_2 \text{ stable matchings,} \quad & \mu_1 \succeq_{\mathcal{M}} \mu_2 \Leftrightarrow \forall m \in \mathcal{M}, \mu_1(m) \succeq_m \mu_2(m) \\ & \mu_1 \succeq_{\mathcal{W}} \mu_2 \Leftrightarrow \forall w \in \mathcal{W}, \mu_1(w) \succeq_w \mu_2(w) \end{aligned}$$

We write $\mu_1 \succ_{\mathcal{W}} \mu_2$ when $\mu_1 \succeq_{\mathcal{W}} \mu_2$ and $\mu_1 \neq \mu_2$, that is when all women weakly prefer μ_1 to μ_2 and some women strictly prefer μ_1 to μ_2. Symmetrically, we write $\mu_2 \succ_{\mathcal{M}} \mu_1$. An important property of stable matchings is that given two stable matchings μ_1 and μ_2, if $\mu_1 \succ_{\mathcal{M}} \mu_2$ then $\mu_2 \succ_{\mathcal{W}} \mu_1$, which proves that both $\succ_{\mathcal{M}}$ and $\succ_{\mathcal{W}}$ define the same ordering (but reversed). The structure of the set of stable matchings was studied by Knuth and Conway [Knu76; Knu97]: with the partial order $\succ_{\mathcal{W}}$, the set stable matching is a distributive lattice.

Definition 2.7. For each person p, we define $\min_p(a_1, a_2)$ and $\max_p(a_1, a_2)$.

$$\min_p(a_1, a_2) = \begin{cases} a_2 & \text{if } a_1 \succ_p a_2 \\ a_1 & \text{otherwise.} \end{cases} \qquad \max_p(a_1, a_2) = \begin{cases} a_1 & \text{if } a_1 \succ_p a_2 \\ a_2 & \text{otherwise.} \end{cases}$$

Given μ_1 and μ_2, we define the "join" $\mu_1 \vee \mu_2$ and the "meet" $\mu_1 \wedge \mu_2$.

$$(\mu_1 \vee \mu_2)(p) = \begin{cases} \min_p(\mu_1(p), \mu_2(p)) & \text{if } p \in \mathcal{M} \\ \max_p(\mu_1(p), \mu_2(p)) & \text{if } p \in \mathcal{W} \end{cases}$$

$$(\mu_1 \wedge \mu_2)(p) = \begin{cases} \max_p(\mu_1(p), \mu_2(p)) & \text{if } p \in \mathcal{M} \\ \min_p(\mu_1(p), \mu_2(p)) & \text{if } p \in \mathcal{W} \end{cases}$$

Observe that operations $\vee$ and $\wedge$ distribute over one another.

Definition 2.7 only uses the fact that μ_1 and μ_2 are functions from $\mathcal{M} \cup \mathcal{W}$ to itself. An important property is that when μ_1 and μ_2 are stable matchings, then both the meet and join are matchings, and those matchings are stable.

Theorem 2.8. *The set of stable matching with the ordering $\succ_{\mathcal{W}}$ is a distributive lattice. More precisely, if μ_1 and μ_2 are stable matchings, then both $\mu_1 \vee \mu_2$ and $\mu_1 \wedge \mu_2$ are stable matchings.*

Proof. We show that $\mu = \mu_1 \vee \mu_2$ is a stable matching. The same proof swapping the roles of men and women proves that $\mu_1 \wedge \mu_2$ is a stable matching.

For the sake of contradiction, assume that there are two women w_1, w_2 and a man m such that $\mu_1(w_1) = \mu(w_1) = m = \mu(w_2) = \mu_2(w_2)$. If $w_1 \succ_m w_2$ then (m, w_1) blocks μ_2, and otherwise (m, w_2) blocks μ_1. Thus, μ is a matching.

Still for the sake of contradiction, assume that there is a pair (m, w) blocking μ. Then we have $w \succ_m \min_m(\mu_1(m), \mu_2(m))$ and $m \succ_w \max_w(\mu_1(w), \mu_2(w))$. If $\mu_1(m) \succ_m \mu_2(m)$ the pair (m, w) blocks μ_2, and if $\mu_2(m) \succ_m \mu_1(m)$ the pair (m, w) blocks μ_1. Thus μ is stable. □

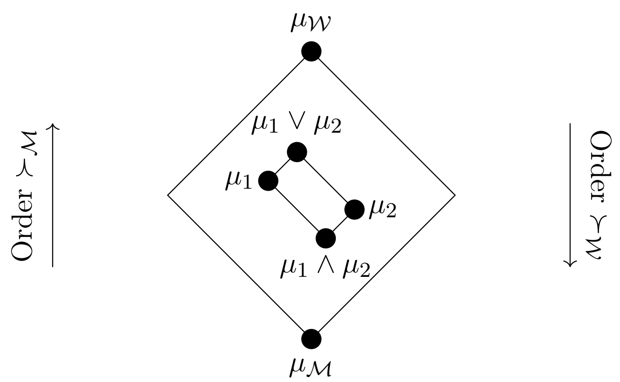


Figure 2.1. Hasse diagram of the lattice of stable matchings

Lemma 2.9. *Given two stable matchings μ_1 and μ_2 such that $\mu_1 \succ_{\mathcal{W}} \mu_2$, if there is a stable matching such that $\mu_1(w) \succ_w \mu(w) \succ_w \mu_2(w)$ for some woman w, then there is an intermediate stable matching μ' such that $\mu'(w) = \mu(w)$ and $\mu_1 \succ_{\mathcal{W}} \mu' \succ_{\mathcal{W}} \mu_2$.*

Proof. Define $\mu' = (\mu \wedge \mu_1) \vee \mu_2$. Because $\wedge$ distributes over $\vee$, we have $\mu' = (\mu \vee \mu_2) \wedge (\mu_1 \vee \mu_2) = (\mu \vee \mu_2) \wedge \mu_1$, thus $\mu_1 \succeq_{\mathcal{W}} \mu' \succeq_{\mathcal{W}} \mu_2$. By construction $\mu'(w) = \mu(w)$, thus $\mu_1 \neq \mu' \neq \mu_2$. □

Definition 2.10. Consider a finite distributive lattice $(\mathcal{L}, \succ)$.

- We say that an element x *covers* an element y and write $x \cdot\!\!\succ y$, if $x \succ y$ and there is no z such that $x \succ z \succ y$.
- An element x is *join-irreducible* if it is not the join of a set of other elements (the bottom element being the join of zero elements). An element x is join-irreducible if and only if it covers a unique element y (x is $\vee$-irreducible $\Leftrightarrow \exists! y,\ x \cdot\!\!\succ y$).
- An element x is *meet-irreducible* if it is not the meet of a set of other elements (the top element being the meet of zero elements). An element x is meet-irreducible it is covered by a unique element y (x is $\wedge$-irreducible $\Leftrightarrow \exists! y,\ y \cdot\!\!\succ x$).

2.4.1 Stable permutations

In this subsection, we will deviate from the classical presentation of the lattice of stable matchings, and define stable permutations (also called stable partitions [Tan91] or half-matchings [BCF08]), that have been introduced for the more general problem of stable roommates and generalize stable matchings. The structure of the lattice of stable matchings will be a corollary of Theorems 2.14, 2.15 and 2.16. This alternative definition has the advantage of keeping the symmetry between men and women, which is not the case using the classical definition of exposed rotation (see [GI89] for a nice survey).

Definition 2.11 (Stable permutation)**.** A *permutation* is a bijection $\sigma : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W}$, where the image of each man m is a woman or himself ($\sigma(m) \in \mathcal{W} \cup \{m\}$), and where the image of each woman w is a man or herself ($\sigma(w) \in \mathcal{M} \cup \{w\}$). For each person p, we use the term *successor* for the image $\sigma(p)$ and the term *predecessor* for the preimage $\sigma^{-1}(p)$.

A man-woman pair (m, w) is blocking permutation σ if m prefers w to his predecessor $\sigma^{-1}(m)$ and w prefers m to her predecessor $\sigma^{-1}(w)$. Abusing notations, a person x is matched with an unacceptable partner if the pair (x, x) is blocking. A permutation σ is *stable* if it has no blocking pair and each person prefer their successor to their predecessor.

Lemma 2.12. *A matching μ is stable if and only if it is stable as a permutation.*

Proof. When a permutation is a matching, Definition 2.11 coincides with Definition 2.20. □

Decomposition in cycles Every a permutation σ can be decomposed into a collection of disjoint cycles. A 1-cycle corresponds to a single person (p such that $\sigma(p) = p$), and a 2-cycle corresponds to a couple (m and w such that $\sigma(m) = w$ and $\sigma(w) = m$).

Definition 2.13. We denote $C(\sigma)$ the number of cycles of length > 2 in a permutation σ.

Theorem 2.14. *We say that σ' is a sub-permutation of σ if for each person p we have:*

$$\binom{\sigma'(p)}{\sigma'^{-1}(p)} \in \left\{ \binom{\sigma(p)}{\sigma(p)}, \binom{\sigma(p)}{\sigma^{-1}(p)}, \binom{\sigma^{-1}(p)}{\sigma^{-1}(p)} \right\}$$

Every permutation σ has exactly $3^{C(\sigma)}$ sub-permutations, and $2^{C(\sigma)}$ of them are matchings. Moreover, if permutation σ is stable then all its sub-permutations are also stable.

Proof. For each cycle with a support $S \subseteq \mathcal{M} \cup \mathcal{W}$ of size > 2, either $\sigma'|_S = \sigma|_S$, or $\sigma'|_S$ is one of the two matchings induced by $\sigma|_S$. Figure 2.3 illustrates the 9 sub-permutations (4 of them being matchings) of a permutation with 2 cycles of length 4. If a pair (m, w) blocks one of the sub-permutation σ', it also blocks σ. □

In particular, given a permutation σ, the restrictions $\sigma|_{\mathcal{M}}$ and $\sigma|_{\mathcal{W}}$ define two matchings μ_1 and μ_2. Conversely, given two matching μ_1 and μ_2, one can define a permutation σ, such that $\sigma|_{\mathcal{M}} = \mu_1|_{\mathcal{M}}$ and $\sigma|_{\mathcal{W}} = \mu_2|_{\mathcal{W}}$. In such situations, we will write $\sigma = \mu_1|_{\mathcal{M}} \sqcup \mu_2|_{\mathcal{W}}$, and Theorem 2.14 to show that if σ is stable then both μ_1 and μ_2 are stable matchings.

Lattice of stable permutations. Lemma 2.12 and Theorem 2.14 show that each stable permutation σ induces $2^{C(\sigma)}$ stable matchings. We will show that stable permutations form a lattice, for

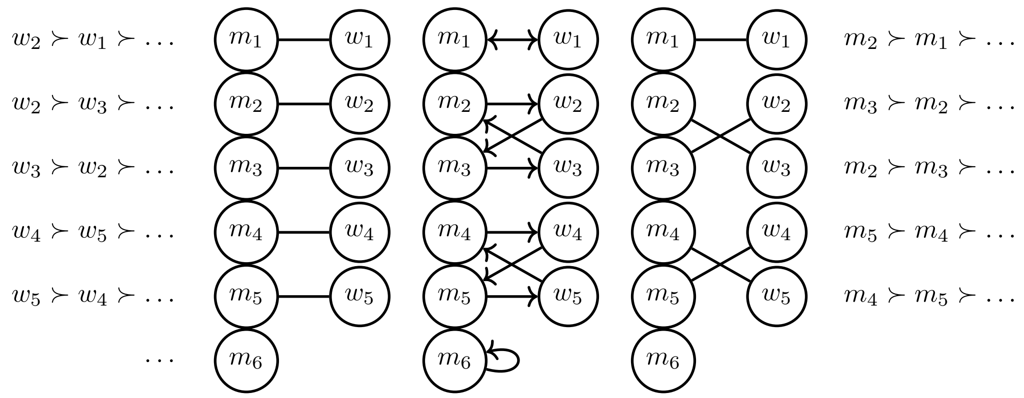


Figure 2.2. Example of stable permutation σ, between two stable matchings.

which the set of stable matchings is a sub-lattice. We start by defining a partial order over stable permutations, which is compatible with the partial order over stable matchings.

$$\sigma_1 \succeq_{\mathcal{M}} \sigma_2 \;\Leftrightarrow\; \forall m \in \mathcal{M},\; \sigma_1(m) \succeq_m \sigma_2(m) \text{ and } \sigma_1^{-1}(m) \succeq_m \sigma_2^{-1}(m)$$
$$\sigma_1 \succeq_{\mathcal{W}} \sigma_2 \;\Leftrightarrow\; \forall w \in \mathcal{W},\; \sigma_1(w) \succeq_w \sigma_2(w) \text{ and } \sigma_1^{-1}(w) \succeq_w \sigma_2^{-1}(w)$$

Once again, we can show that given two stable permutations σ_1 and σ_2, we have $\sigma_1 \succeq_{\mathcal{M}} \sigma_2$ if and only if $\sigma_2 \succeq_{\mathcal{W}} \sigma_1$, which proves that $\succeq_{\mathcal{M}}$ and $\succeq_{\mathcal{W}}$ induce the same ordering (but reversed). Notice that Definition 2.7 can be used to compute the meet and join of two permutations.

Theorem 2.15. *The set of stable permutations with the ordering $\succ_{\mathcal{W}}$ is a distributive lattice. More precisely, if σ_1 and σ_2 are stable permutations, then both $\sigma_1 \vee \sigma_2$ and $\sigma_1 \wedge \sigma_2$ are stable permutations.*

Proof. The proof is identical to the one of Theorem 2.8 □

Relations between permutations and matchings. So far, we gave several results to manipulate and combine stable permutations, but we did not prove the existence of stable permutations that are not matchings. Theorem 2.16 show that stable permutations with one cycle of length > 2 exactly corresponds to the edges of the Hasse diagram of the lattice of stable matchings (covering relation $\mu_1 \cdot\!\!\succ_{\mathcal{M}} \mu_2$).

Theorem 2.16. *Let μ_1 and μ_2 be matchings that induce a permutation $\sigma = \mu_1|_{\mathcal{M}} \sqcup \mu_2|_{\mathcal{W}}$. Then, the following two conditions are equivalent:*

- *Both μ_1 and μ_2 are stable, and $\mu_1 \cdot\!\!\succ_{\mathcal{M}} \mu_2$.*
- *Permutation σ is stable, and $C(\sigma) = 1$.*

The proof of Theorem 2.16 is technical, and relies on Lemmas 2.17 and 2.18.

Lemma 2.17. *Let μ_1 and μ_2 be two stable matchings such that permutation $\sigma = \mu_1|_{\mathcal{M}} \sqcup \mu_2|_{\mathcal{W}}$ is not stable. Then one of the following three properties hold:*

(1) One person p prefers their predecessor to their successor, that is $\sigma^{-1}(p) \succ_p \sigma(p)$.

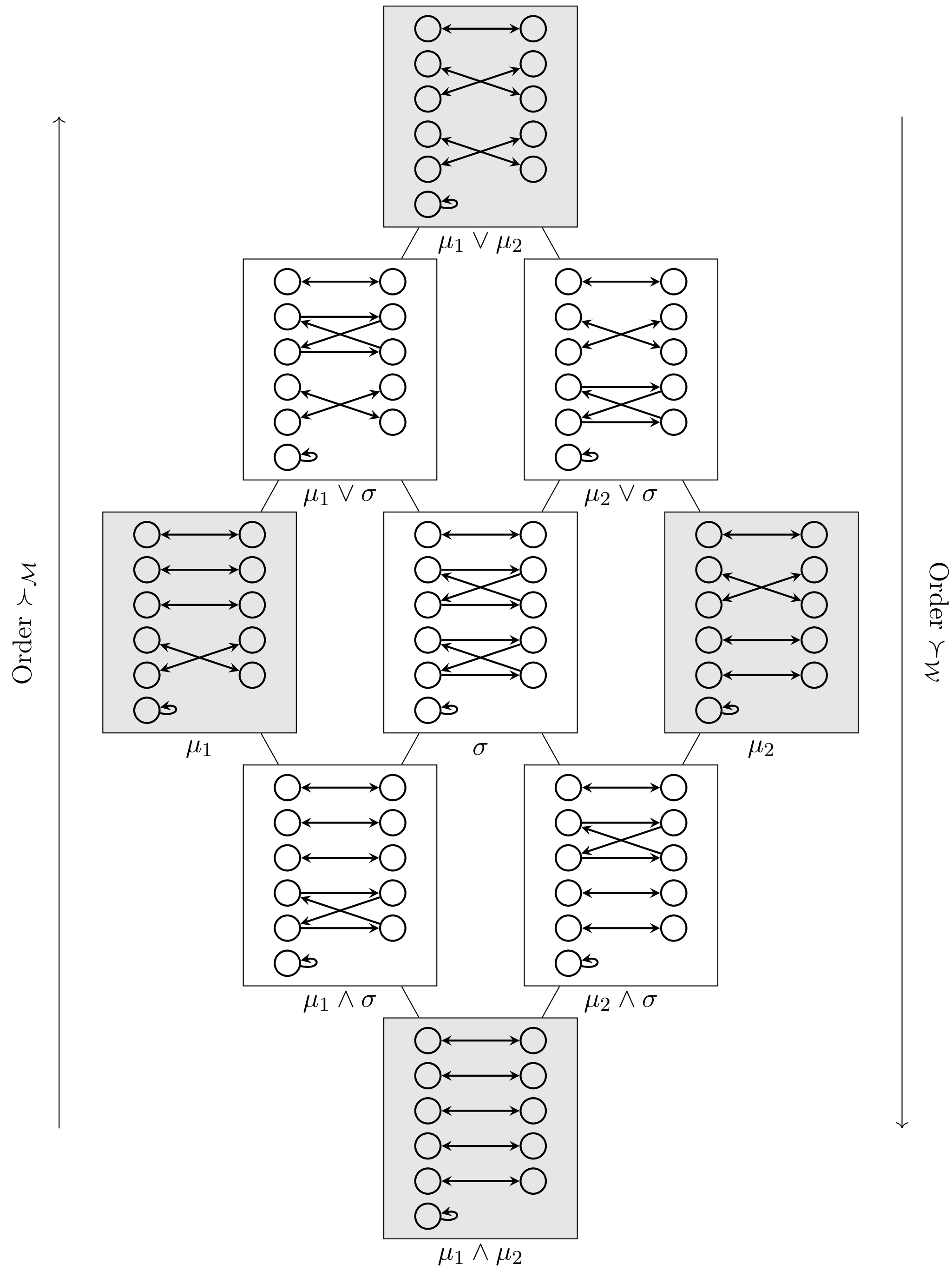


Figure 2.3. Lattice structure of stable permutations. The central permutation σ has 2 cycles of length 4, and induces $3^{C(\sigma)} = 9$ sub-permutations, with $2^{C(\sigma)} = 4$ of them being matchings (in gray).

(2) There is a stable matching μ such that $\sigma(p) \succ_p \mu(p) \succ_p \sigma^{-1}(p)$ for some p.

(3) There are two matchings μ and μ', where μ is stable and μ' is not, and such that $\{\mu(p), \mu'(p)\} = \{\mu_1(p), \mu_2(p)\}$ for all p.

Proof. If (1) holds the proof is finished. Thus we assume that each person p prefer their successor to their predecessor, that is $\sigma(p) \succeq_p \sigma^{-1}(p)$. Consider that each person p truncates their preference list and declare unacceptable persons to whom they prefer $\sigma^{-1}(p)$. Repeatedly using Lemma 2.6, we show that each hypothesis/conclusion holds if and only if it holds under the original preferences. For the rest of the proof we work with truncated preferences.

Consider the extremal stable matchings $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{W}}$. We have $\mu_{\mathcal{M}}(m) = \sigma(m) = \mu_1(m)$ for each man m, and $\mu_{\mathcal{W}}(w) = \sigma(w) = \mu_2(w)$ for each woman w. Because σ is not stable, there is a pair (m, w) such that $m \succ_w \sigma^{-1}(w)$ and $w \succ_m \sigma^{-1}(m)$. By stability of $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{W}}$, we have $\mu_{\mathcal{M}}(m) \succ_m w \succ_m \mu_{\mathcal{W}}(m)$ and $\mu_{\mathcal{W}}(w) \succ_w w \succ_w \mu_{\mathcal{M}}(w)$. We use Algorithm 2.2 to enumerate stable partners of w. Consider the matching μ obtained at the end of Algorithm 2.2 (when w accepts the last proposal, from $\mu_{\mathcal{W}}(w)$). Using Theorem 2.5, matching μ is stable.

- If there is a person p such that $\sigma(p) \succ_p \mu(p) \succ_p \sigma^{-1}(p)$, then (2) holds.
- If $\mu(m) = \mu_{\mathcal{W}}(m)$, then m proposed to w during the execution of Algorithm 2.2. Thus w accepted at least one proposal before the last proposal from $\mu_{\mathcal{W}}(w)$, which proves that w has an intermediate stable partner, which proves (2).
- Otherwise, $\mu(m) = \mu_{\mathcal{M}}(m)$ and $\mu(p) \in \{\mu_{\mathcal{M}}(p), \mu_{\mathcal{W}}(p)\}$ for each person p. We consider the matching μ', such that $\mu'(p) \in \{\mu_{\mathcal{M}}(p), \mu_{\mathcal{W}}(p)\} \setminus \{\mu(p)\}$ for each person p. Then (m, w) blocks μ', which proves that (3) holds.

□

Lemma 2.18. *A permutation σ is stable if and only if the following conditions hold:*

- *Every person p prefers their successor to their predecessor, that is $\sigma(p) \succeq_p \sigma^{-1}(p)$*
- *If μ is a matching such that $\sigma(p) \succ_p \mu(p) \succ_p \sigma^{-1}(p)$ for some p, then μ is not stable.*
- *If μ is a matching such that $\mu(p) \in \{\sigma(p), \sigma^{-1}(p)\}$ for all p, then μ is stable.*

Proof. Assuming that all three conditions hold, the contrapositive of Lemma 2.17 proves that σ is stable. Conversely, assuming that σ is a stable permutation, then the first condition holds by definition and the second holds because no pair can block μ without blocking σ. For the sake of contradiction, assume that there is a stable matching μ such that $\sigma(p) \succ_p \mu(p) \succ_p \sigma^{-1}(p)$ for some person p. We are going to show that the pair (p, p') with $p' = \mu(p)$ blocks permutation σ. We already know that $p' \succ_p \sigma^{-1}(p)$, and we are going to show that $p \succ_{p'} \sigma^{-1}(p')$. Define a sequence of persons $(p_k)_{k\geq 0}$ with $p_0 = p$ and $p_{k+1} = \mu(\sigma(p_k))$ for all $k \geq 0$. By induction on $k \geq 0$ we show that

$$\forall k \geq 0, \quad p_{k+1} \succ_{\sigma(p_k)} p_k \qquad \text{(because } \mu \text{ is stable and } \sigma(p_k) \succ_{p_k} \mu(p_k))$$
$$\sigma(p_{k+1}) \succ_{p_{k+1}} \sigma(p_k) \qquad \text{(because } \sigma \text{ is stable and } p_{k+1} \succ_{\sigma(p_k)} p_k)$$

Because both σ and μ are permutations, the sequence is periodic and $p = p_{k+1}$ for some $k \geq 0$. For such k, we have $p' = \sigma(p_k)$ and $p \succ_{p'} \sigma^{-1}(p')$, which contradicts the fact that σ is stable, which proves that the third condition holds. □

Proof of Theorem 2.16. First, we assume that σ is stable and that $C(\sigma) = 1$. Lemma 2.18 shows that μ_1 and μ_2 are stable, and that $\mu_1 \succeq_{\mathcal{M}} \mu_2$. For the sake of contradiction, assume that there is a stable matching μ such that $\mu_1 \succ_{\mathcal{M}} \mu \succ_{\mathcal{M}} \mu_2$. Using Lemma 2.18, μ matches each person p with $\mu_1(p)$ or $\mu_2(p)$. Because there is exactly $C(\sigma) = 1$ cycle, μ must be either equal to μ_1 or μ_2, which is a contradiction.

We now assume that μ_1 and μ_2 are stable, and that $\mu_1 \mathrel{\cdot\!\succ}_{\mathcal{M}} \mu_2$. For the sake of contradiction, assume that σ is not stable and use Lemma 2.17. If (1) holds, we have a contradiction with $\mu_1 \succ_{\mathcal{M}} \mu_2$. If (3) holds there is an intermediate matching $\mu_1 \succ_{\mathcal{M}} \mu \succ_{\mathcal{M}} \mu_2$. If (2) holds, we use Lemma 2.9 and show that there is also an intermediate matching. Finally, if $C(\sigma) > 1$, one can build a matching μ distinct from μ_1 and μ_2, such that $\mu(p) \in \{\sigma(p), \sigma^{-1}(p)\}$ for each person p. Lemma 2.18 proves that μ is stable, and it is intermediate by construction. Having an intermediate matching contradicts $\mu_1 \mathrel{\cdot\!\succ}_{\mathcal{M}} \mu_2$, which concludes the proof. ☐

2.4.2 Algorithm to compute stable permutations

Equipped with our understanding of the relation between stable matchings and stable permutations, we can now design an algorithm to compute stable permutations.

Algorithm 2.3 Computing a stable permutation.

Input: Preferences of men $(\succ_m)_{m \in M}$ and of women $(\succ_w)_{w \in \mathcal{W}}$. Matching μ.
Initialize the function $\tau \leftarrow \mu$.
For each man m**, do**
  If there is a woman w such that $\mu(m) \succ_m w$ and $m \succ_w \mu(w)$**, then**
    Let w be m's favourite partner among such women, and set $\tau(m) \leftarrow w$.
For each person p who does not belong to a cycle of τ**, do**
  Set $\tau(p) \leftarrow \mu(p)$.
Output: Permutation $\sigma = \tau^{-1}$.

Theorem 2.19. *Given a stable matching μ, Algorithm 2.3 outputs a stable permutation $\sigma = \bigwedge\{\sigma'$ stable permutation such that $\sigma'|_{\mathcal{M}} = \mu|_{\mathcal{M}}\}$.*

Proof. By construction, Algorithm 2.3 outputs a permutation σ such that $\sigma|_{\mathcal{M}} = \mu|_{\mathcal{M}}$. For the sake of contradiction, assume that σ is not stable: there is a blocking pair (m, w), where $m \succ_w \sigma^{-1}(w) = \tau(w) = \mu(w)$ and $w \succ_m \sigma^{-1}(m) = \tau(m)$. By stability of μ, we have $\mu(m) \succ_m w$. This contradicts the fact that $\tau(m)$ is m's favourite partner among such women. Thus σ is stable.

Let σ' be a stable permutation such that $\sigma'|_{\mathcal{M}} = \mu|_{\mathcal{M}}$. For each man m such that $\sigma'^{-1}(m) \neq \mu(m)$, we must have $\tau(m) = \sigma'^{-1}(m)$. Because m belongs to a cycle of length > 2, he belongs to a cycle in τ, and thus $\sigma^{-1}(m) = \sigma'^{-1}(m)$. ☐

2.4.3 Rotations

Stable permutations are closely related to the (more classical) concept of rotation, that was introduced by Irving, Leather and Gusfield [IL86; GI89]. We view a *rotation* as a simple directed cycle r in the complete bipartite graph over $\mathcal{M} \cup \mathcal{W}$. When a person $x \in \mathcal{M} \cup \mathcal{W}$ belongs to the cycle, we write $x \in r$, denote $r(x)$ a successor and $r^{-1}(x)$ a predecessor. In a stable matching μ_1, rotation r is exposed and women-improving if for all man m, $r(m) = \mu_1(m)$ and $r^{-1}(m)$ is m's favourite woman among women w to whom he prefers his wife $(\mu_1(m) \succ_m w)$, and who prefer m to their husband ($m \succ_w \mu_1(w)$). Eliminating rotation r in matching μ_1 creates a new stable matching μ_2; we have $\mu_2 \succeq_{\mathcal{W}} \mu_1$.

$$\forall m \in M,\ \mu_2(m) = \begin{cases} r^{-1}(m) & \text{if } m \in r \\ \mu_1(m) & \text{if } m \notin r \end{cases} \qquad \forall w \in \mathcal{W},\ \mu_2(w) = \begin{cases} r(w) & \text{if } w \in r \\ \mu_1(w) & \text{if } w \notin r \end{cases}$$

Symmetrically, rotation r is exposed and men-improving in stable matching μ_2. Eliminating r in μ_2 creates stable matching μ_1; we have $\mu_1 \succeq_{\mathcal{M}} \mu_2$.

Given a matching μ and a set of disjoint rotations R, one can build the associated permutation σ_R. If $C(\sigma)$ is the number of cycles of length > 2 in σ, we have $C(\sigma_R) = |R|$. A corollary of Theorem 2.19 is that σ_R is stable if and only if μ is stable and every rotation from R is exposed and women-improving.

$$\sigma_R : \begin{cases} m \mapsto \mu(m) & \text{if } m \in \mathcal{M} \\ w \mapsto \mu(w) & \text{if } w \in \mathcal{W} \text{ and } w \notin r \text{ for all } r \in R \\ w \mapsto r(w) & \text{if } w \in \mathcal{W} \text{ and } w \in r \text{ for some } r \in R \end{cases}$$

2.5 Birkhoff's Representation Theorem

Birkhoff's Representation Theorem [Bir37] states that any finite distributive lattice is isomorphic to the lattice of lower sets of the partial order of its join-irreducible elements. Irving, Leather and Gusfield [IL86; Gus87] proved that rotations are isomorphic to the join-irreducible stable matchings, and gave efficient algorithms to compute the partially ordered set of rotations. For the statements and proofs of those results, we refer the reader to the excellent book of Gusfield and Irving [GI89], even though we believe that the proof of most of these results could be derived from Theorems 2.14, 2.15 and 2.16.

Because a theorem is best understood with a nice illustration, we implemented (in C++) the algorithms of Irving, Leather and Gusfield, and generated several animations (in HTML, SVG and Javascript) to visualize the correspondence between stable matchings and closed sets of rotations. Figure 2.4 is a screenshot of one of those animations. Implementations are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/Lattice

2.6 Many-to-one matching markets

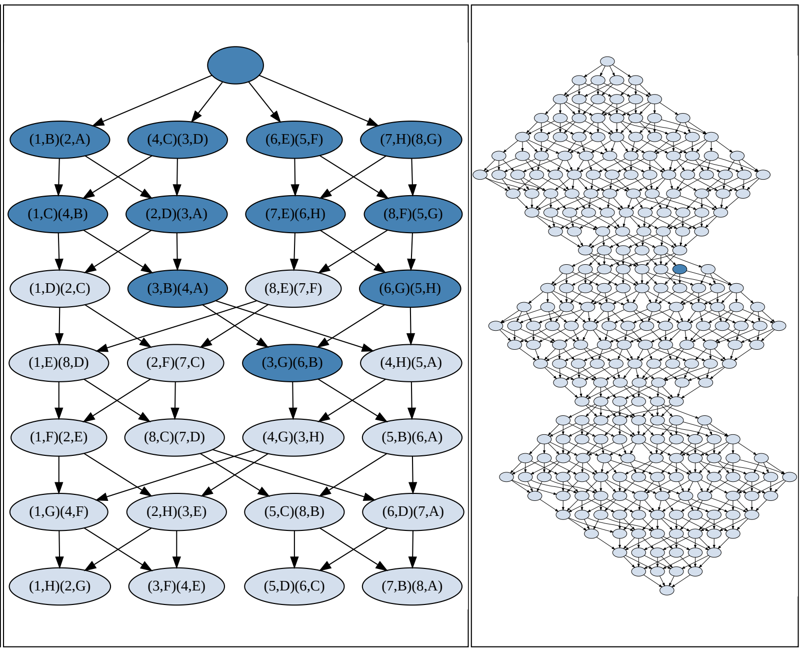
The original motivation of Gale and Shapley [GS62] when introducing stable marriages was the college admissions matching markets. Let $\mathcal{S} = \{s_1 \dots s_S\}$ be a set of S students, and let $\mathcal{C} = \{c_1, \dots, c_C\}$ be a set of C colleges. Each agent has a capacity defined by $b : \mathcal{S} \cup \mathcal{C} \to \mathbb{N}$, where $b(s) = 1$ for each student s. Similarly to the one-to-one case, each agent gives a strict ordering on itself and the members of the opposite side.

Definition 2.20 (Stable b-matching)**.** Formally, a bipartite b-matching is a function $\nu : \mathcal{S} \times \mathcal{C} \to \mathbb{N}$, where each student s is paired with at most one college ($\sum_c \nu(s,c) \leq 1$), and where each college c is matched with at most $b(c)$ students ($\sum_s \nu(s,c) \leq b(c)$). For convenience, we denote $\nu(a)$ the multi-set of agents with whom a is matched, possibly including multiple copies of a itself in order to reach the capacity $|\nu(a)| = b(a)$.

A student-college pair (s,c) is blocking if $s \succ_c \min_c \nu(c)$ and $c \succ_s \min_s \nu(s)$, where $\min_a$ denotes the worst partner of a within a set. Abusing notations, observe that ν matches a agent

1	2	3	4	5	6	7	8
A	B	C	D	E	F	G	H
B	A	D	C	F	E	H	G
C	D	A	B	G	H	E	F
D	C	B	A	H	G	F	E
E	F	G	H	A	B	C	D
F	E	H	G	B	A	D	C
G	H	E	F	C	D	A	B
H	G	F	E	D	C	B	A

A	B	C	D	E	F	G	H
8	7	6	5	4	3	2	1
7	8	5	6	3	4	1	2
6	5	8	7	2	1	4	3
5	6	7	8	1	2	3	4
4	3	2	1	8	7	6	5
3	4	1	2	7	8	5	6
2	1	4	3	6	5	8	7
1	2	3	4	5	6	7	8



Preference lists Closed sets of rotations Lattice of stable matchings

Figure 2.4. Birkhoff's Theorem states that the set of stable matchings is in correspondence with upward closed sets of rotations (if one rotation belong to the set, all its ancestors also do). The animation (with clickable matchings/rotations) is available at the following address: https://www.irif.fr/~mauras/stablematchings/7/

a with an unacceptable partner when a would prefer to remain single, that is when the pair (a,a) is blocking. A b-matching with no blocking pair is stable.

Observe that the definitions generalize the one-to-one setting (and would also work in a many-to-many setting). A simple reduction allows us to transpose all the results from the one-to-one setting.

Theorem 2.21. *Define the sets $\mathcal{M} = \{m_i \mid i \in [S]\}$ and $\mathcal{W} = \{w_j^k \mid j \in [C], k \in [b(c_j)]\}$, where colleges are duplicated to account for capacities. We build the preference list of each man/women using the corresponding preferences of students/colleges, replacing s_i by m_i and replacing c_j by $w_j^1 \succ \cdots \succ w_j^{b(c_j)}$.*

- *Given a matching $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W}$, we build a b-matching ν.*

$$\nu : \begin{cases} \mathcal{S} \times \mathcal{C} & \to \quad \mathbb{N} \\ (s_i, c_j) & \mapsto \quad \sum_k \mathbb{1}[\mu(m_i) = w_j^k] \end{cases}$$

- *Given a b-matching ν, we build a matching $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W}$ such that $\mu(m_i) = w_j^k$ if $\nu(s_i, c_j) = 1$ and s_i is the k-th choice of c_j in $\nu(c_j)$; and m_i otherwise.*

Then a matching μ is stable if and only if the corresponding b-matching ν is stable.

Proof. By construction, a person is matched with a non-acceptable partner in μ if and only if the corresponding agent is matched with an acceptable partner in ν. If a pair (m_i, w_j^k) blocks μ, then the pair (s_i, c_j) will block ν. If a pair (s_i, c_j) blocks ν, then the pair $(m_i, w_j^{b(c_j)})$ will block μ. □

2.7 Incentive compatibility

A procedure is *Dominant Strategy Incentive Compatible* (DSIC, also called strategy-proof or truthful) if truth-telling is a (weakly) dominant strategy; that is if for every agent, reporting one's true preference list is weakly better than lying, whichever the other agents' preferences are. When misreporting their preferences, we will assume that agents are allowed to declare some partners unacceptable, and we refer the reader to [TST01] for the case where agents are only allowed to reorder their acceptable partners.

Dubins and Freedman [DF81] showed that the men-proposing deferred acceptance algorithm is strategy-proof against coalitions of men, and symmetrically that the women-proposing deferred acceptance algorithm is strategy-proof against coalitions of women. Interestingly, Roth [Rot82] showed that even though deferred acceptance is truthful for the proposing side, no procedure which selects a stable matching can be simultaneously truthful for men and women (see Figure 2.5 for a counter example).

Gale and Sotomayor [GS85a] studied the extent to which women can manipulate in the men-proposing deferred acceptance procedure, and showed that each woman can ensure that she is matched with her best stable partner. Demange, Gale and Sotomayor [DGS87] investigated the strategy-proofness of stable matching procedures against general coalitions containing men and women, and proved Theorem 2.22 which generalizes the results from [DF81] and [GS85a]. The proof if based on Lemma 2.23, credited to J.S. Hwang in a paper of Gale and Sotomayor [GS85b].

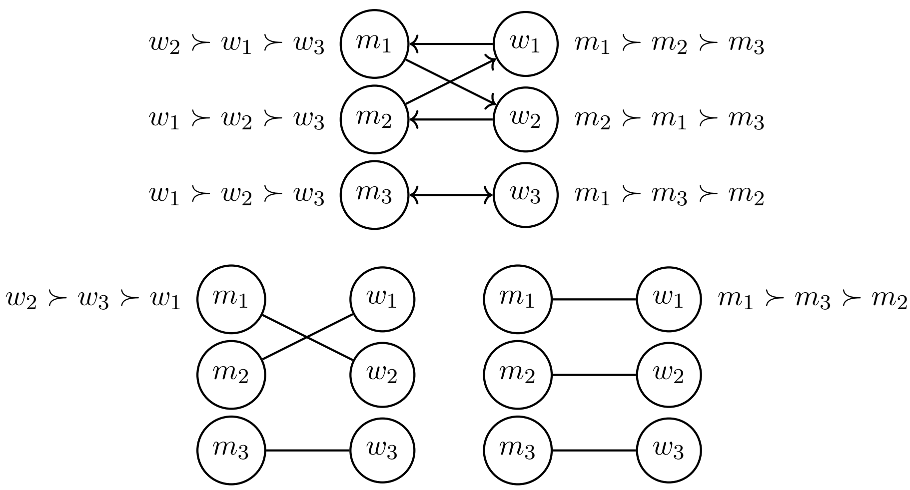


Figure 2.5. No procedure which selects a stable matching can be simultaneously truthful for men and women. The instance with original preferences has exactly two stable matchings, $\mu_{\mathcal{M}}$ and $\mu_{\mathcal{W}}$, that we represent using the corresponding stable permutation. If the procedure selects $\mu_{\mathcal{W}}$, then m_1 can misreport his preference to make sure that the only stable matching is $\mu_{\mathcal{M}}$; whereas if the procedure selects $\mu_{\mathcal{M}}$, then w_1 can misreport her preferences to make sure that the only stable matching is $\mu_{\mathcal{W}}$.

Theorem 2.22 (Adapted from [DGS87])**.** *Given an instance of stable matching, let* $C \subseteq \mathcal{M} \cup \mathcal{W}$ *be a coalition of men and women who misreport their preferences. For every matching* μ *stable under the misreported preferences, there is a matching* μ' *stable under the original preferences, such that* $\mu'(p) \succeq \mu(p)$ *for some* $p \in C$*.*

Proof. For the sake of contradiction, assume that there is a matching μ stable under misreported preferences of a coalition C, such that every person $p \in C$ prefer μ to any matching μ' stable under the original preferences. Then we have:

$$\forall m \in C \cap \mathcal{M}, \quad \mu(m) \succ_m \mu_{\mathcal{M}}(m) \qquad \text{and} \qquad \forall w \in C \cap \mathcal{W}, \quad \mu(w) \succ_w \mu_{\mathcal{W}}(w)$$

In particular, agents in the coalition strictly improved their outcome and thus are matched with acceptable partners. Agents not in the coalition reported their true preferences and thus are matched with acceptable partners. Without loss of generality, we assume that $C \cap \mathcal{M} \neq \emptyset$ (otherwise, swap the roles of men and women). Using Lemma 2.23, there is a man $m \in \mathcal{M} \setminus C$ and a woman $w \in \mu(\mathcal{M} \cap C)$ such that (m, w) blocks μ under the original preferences. Considering $m' = \mu(w) \in C$, we have $w = \mu(m') \succ_{m'} \mu_{\mathcal{M}}(m')$, and thus $\mu_{\mathcal{M}}(w) \succ_w \mu(w) = m'$ by stability of $\mu_{\mathcal{M}}$, which proves that $w \in \mathcal{W} \setminus C$. Because neither m nor w belong to the coalition, they also block μ under misreported preferences, which is a contradiction. □

Lemma 2.23 (Adapted from [GS85b])**.** *Let* μ *be a matching where each person is matched with an acceptable partner, and let* $S = \{m \in \mathcal{M} \mid \mu(m) \succ_m \mu_{\mathcal{M}}(m)\}$ *be the set of men who prefer* μ *to* $\mu_{\mathcal{M}}$*. If* S *is non-empty, then there is a man* $m \in \mathcal{M} \setminus S$ *and a woman in* $\mu(S)$ *such that* (m, w) *blocks* μ*.*

Proof. This is a corollary of the proof of Theorem 2.3, when we show that $\mu_{\mathcal{M}}$ matches every man to his favourite stable partner. A case by case analysis shows that the pair (m, w) blocking μ satisfies $m \in \mathcal{M} \setminus S$ and $w \in \mu(S)$. □

Corollary 2.24 (Adapted from [DF81])**.** *The men (resp. women) proposing deferred acceptance procedure is strategy-proof against coalitions of men (resp. women).*

Proof. Assume that a coalition of men C misreport their preferences, such that the men optimal stable matching becomes. Using Theorem 2.22, at least one of them prefers $\mu_{\mathcal{M}}$ to $\mu'_{\mathcal{M}}$, and thus has no (strong) incentive to participate in the coalition. □

Corollary 2.25 (Adapted from [GS85a])**.** *Each person's incentives to manipulate are bounded by the difference between their worst and best stable partner. In particular, if there is a unique stable matching, then any procedure which selects a stable matching is strategy-proof against general coalitions.*

Proof. Theorem 2.22 shows that a strategic agent p who misreport their preferences cannot get a better outcome than their best stable partner p'. Conversely, a strategic agent p can truncate their preference list after their best stable partner p', which leads to a preference profile for which every stable matching matches p and p'. □

A natural question to ask is whether Corollary 2.24 extends to the many-to-one setting where each college corresponds to a coalition. In particular, one might wonder if the college-proposing deferred acceptance procedure is strategy-proof for colleges. The answer has been given by [Rot85], who built an instance (see Figure 2.6) where a college can misreport its preference list to get a strictly better outcome than in the college optimal stable matching. The intuition is that Corollary 2.24 does not preclude the existence of a coalition of men where each man is weakly better and some men are strictly better than in the men-optimal stable matching. The robustness of this phenomenon was later explored by Huang [Hua06].

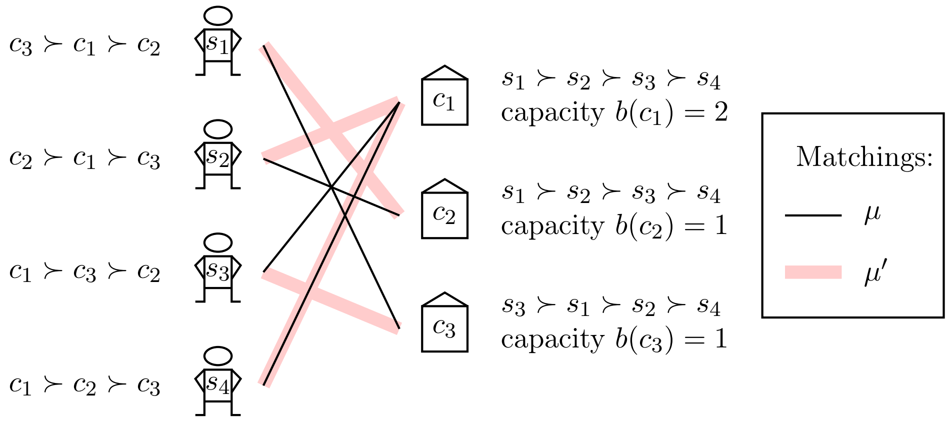


Figure 2.6. The college-proposing deferred acceptance procedure is not strategy-proof for colleges. Observe that all colleges strictly prefer matching μ' to matching μ. Under the original preferences, μ is the unique stable matching. If college c_1 reports the preference list $s_2 \succ s_4 \succ s_1 \succ s_3$, then μ' becomes the college optimal stable matching.

3 | Random Preferences

We consider models where each person's set of acceptable partners is deterministic, and orderings of acceptable partners are drawn independently from a distribution. We say that preferences are *complete* when every partner is acceptable. When unspecified, someone's acceptable partners and/or their ordering is adversarial, chosen by an adversary who knows the input model but does not know the outcome of the random coin flips.

Critchlow, Fligner and Verducci [CFV91] wrote a nice survey on the different random models on ranking that have been proposed in the statistical and psychological literature. They partition existing model in four categories: (1) order statistics models, (2) ranking models induced by pairwise comparisons, (3) ranking models based on distance between permutations and (4) multistage ranking models. We will mainly focus on the class (1), which corresponds to cardinal preferences, and that we will call *utility preferences*. In the matching literature, ordinal preferences have been modeled by *popularity preferences* [IM15; KP09; Ash+21], which belong to both classes (1) and (4). Models discussed in this thesis are summarized in Figure 3.1.

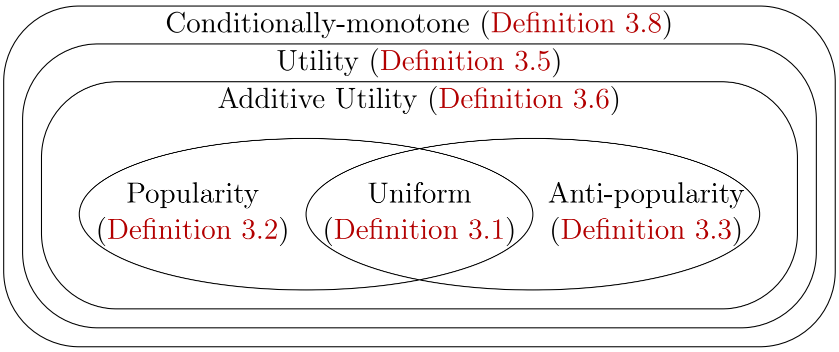


Figure 3.1. Classes of random preferences

3.1 Popularity-based preferences

The first and probably the most studied model of random preferences is the uniform distribution.

Definition 3.1 (Uniform preferences)**.** We say that someone has uniform preferences if they order uniformly at random the agents they declare acceptable.

The *Choice Axiom* is a property introduced by Luce [Luc59; Luc77], which is reminiscent of the irrelevance of independent alternatives in voting theory. It states that when selecting the best agent from a set, the probability to pick one agent over another is independent from the set of agents. This leads to a definition of distribution, where agents are picked from best to worst with probability proportional to a positive value we call *popularity*.

Definition 3.2 (Popularity preferences)**.** Someone has popularity preferences induced by $P : A \to \mathbb{R}_+$ if they give popularity $P(a)$ to each agent $a \in A$, where $P(a) = 0$ means that a is not acceptable. They first build a "popularity" distribution over acceptable partners, where a has probability $P(a)$, scaled such that probabilities sum up to 1. Then they build a preference list by sampling without replacement from this "popularity" distribution: they draws first their favourite partner, then their second favourite, and so on until their least favourite partner.

As stated in [CFV91], popularity preferences are not reversible, in the sense that the mirror ordering of popularity preferences are not popularity preferences, but is a class of distributions we will call anti-popularity preferences.

Definition 3.3 (Anti-popularity preferences)**.** Someone has anti-popularity preferences induced by $P : A \to \mathbb{R}_+$ if they give popularity $P(a)$ to each agent $a \in A$, where $P(a) = 0$ means that a is not acceptable. They first build an "anti-popularity" distribution over acceptable partners, where a has probability $1/P(a)$, scaled such that probabilities sum up to 1. Then they build a preference list from the end, by sampling without replacement from this "anti-popularity" distribution: they draw first their least favourite partner, then their second least, and so on until their favourite partner.

Consider the following popularities

$$P(a_1) = 8, \quad P(a_2) = 1, \quad P(a_3) = 4.$$

Then $a_2 \succ a_1 \succ a_3$ with probability

$$\frac{1}{1+4+8} \cdot \frac{8}{4+8} \cdot \frac{4}{4} \approx 0.05$$

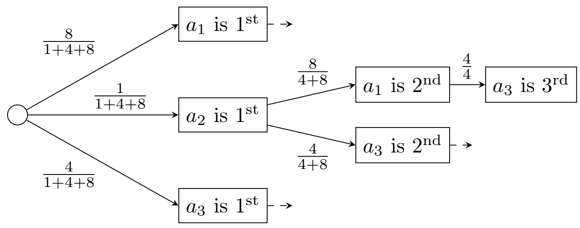


Figure 3.2. Popularity preference distribution

Consider the following popularities

$$P(a_1) = 8, \quad P(a_2) = 1, \quad P(a_3) = 4.$$

Then $a_2 \succ a_1 \succ a_3$ with probability

$$\frac{1/4}{1/1+1/4+1/8} \cdot \frac{1/8}{1/1+1/8} \cdot \frac{1/1}{1/1} \approx 0.02$$

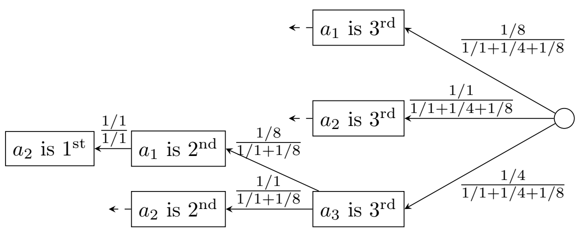


Figure 3.3. Anti-popularity preference distribution

Given a popularity function $P : A \to \mathbb{R}_+$, both the popularity and anti-popularity preferences induce the same probabilities for pairwise comparisons, namely $\mathbb{P}[a_1 \succ a_2] = P(a_1)/(P(a_1)+P(a_2))$. Thus, by linearity of expectation, the expected rank of each agent is the same under popularity and anti-popularity preferences.

The difference between the two distributions comes from rare events. As an example, consider the following situation: n agents $a_1, \dots a_n$ have popularity n, one agent b has popularity $\sqrt{n}$, and

n agents $c_1, \ldots, c_n$ have popularity 1. In both settings, the expected rank of agent b is $1 + n$:

$$\mathbb{E}[\text{rank of } b] = 1 + \sum_{i=1}^{n} \underbrace{\mathbb{P}[a_i \succ b]}_{\frac{n}{n+\sqrt{n}}} + \sum_{i=1}^{n} \underbrace{\mathbb{P}[c_i \succ b]}_{\frac{1}{1+\sqrt{n}}} = 1 + n$$

Agents $a_1, \ldots, a_n$ are popular, and can be ranked unusually low by anti-popularity preferences, thus agent b will likely precede one of them. Agents $c_1, \ldots, c_n$ are unpopular, and can be ranked unusually high by popularity preferences, thus one of them will likely precede agent b. More precisely, we have:

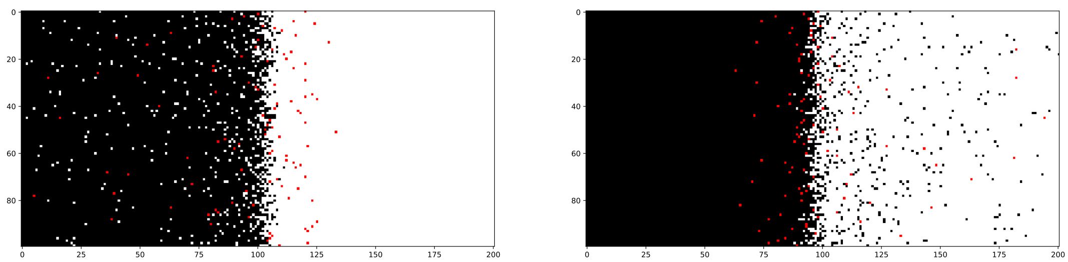
$$\mathbb{P}[b \succ c_1, \ldots, c_n] = \frac{\sqrt{n}}{\sqrt{n} + n \cdot 1} \sim \frac{1}{\sqrt{n}} \qquad \text{(Definition 3.2)}$$

$$\mathbb{P}[a_1, \ldots, a_n \succ b] = \frac{\frac{1}{\sqrt{n}}}{\frac{1}{\sqrt{n}} + n \cdot \frac{1}{n}} \sim \frac{1}{\sqrt{n}} \qquad \text{(Definition 3.3)}$$

Conversely, agents $a_1, \ldots, a_n$ are likely to precede b in the popularity setting, and b is likely to precede $c_1, \ldots, c_n$ in the anti-popularity setting.

$$\mathbb{P}[a_1 \ldots, a_n \succ b] = \prod_{i=1}^{n} \frac{i \cdot n}{i \cdot n + \sqrt{n}} = 1 - \frac{\ln(n)}{\sqrt{n}} - \mathcal{O}(\tfrac{1}{\sqrt{n}}) \qquad \text{(Definition 3.2)}$$

$$\mathbb{P}[b \succ c_1, \ldots, c_n] = \prod_{i=1}^{n} \frac{i \cdot \frac{1}{1}}{i \cdot \frac{1}{1} + \frac{1}{\sqrt{n}}} = 1 - \frac{\ln(n)}{\sqrt{n}} - \mathcal{O}(\tfrac{1}{\sqrt{n}}) \qquad \text{(Definition 3.3)}$$



Popularity (Definition 3.2) Anti-popularity (Definition 3.3)

Figure 3.4. Comparing the popularity and anti-popularity distributions: n agents $a_1, \ldots, a_n$ in black have popularity n, one agent b in red has popularity $\sqrt{n}$, and n agents in white have popularity 1. Each panel plots 100 draws, agents are ordered from left to right.

Lemma 3.4. *Consider a preference distribution $\succ$ over at least 3 acceptable agents. Then $\succ$ is the uniform distribution if and only if it belongs to both the classes of popularity and anti-popularity distributions.*

Proof. Choosing $P : a \mapsto \mathbb{1}[a \text{ is acceptable}]$ yields an incomplete uniform distribution, in both the popularity and anti-popularity settings. Conversely, let a, b and c be three acceptable agents. Both with popularity and anti-popularity preferences, we use pairwise comparisons to compute ratios of popularities:

$$\frac{\mathbb{P}[a \succ b]}{\mathbb{P}[b \succ a]} = \frac{P(a)}{P(b)} \qquad \frac{\mathbb{P}[a \succ c]}{\mathbb{P}[c \succ a]} = \frac{P(a)}{P(c)} \qquad \frac{\mathbb{P}[b \succ c]}{\mathbb{P}[c \succ b]} = \frac{P(b)}{P(c)}.$$

Thus, we can assume that popularity and anti-popularity preferences are induced by the same popularity function P.

$$\frac{P(a)}{P(a)+P(b)+P(c)} \cdot \frac{P(b)}{P(b)+P(c)} = \mathbb{P}[a \succ b \succ c] = \frac{\frac{1}{P(c)}}{\frac{1}{P(a)}+\frac{1}{P(b)}+\frac{1}{P(c)}} \cdot \frac{\frac{1}{P(b)}}{\frac{1}{P(a)}+\frac{1}{P(b)}} \tag{3.1}$$

$$\frac{P(c)}{P(a)+P(b)+P(c)} \cdot \frac{P(b)}{P(b)+P(a)} = \mathbb{P}[c \succ b \succ a] = \frac{\frac{1}{P(a)}}{\frac{1}{P(a)}+\frac{1}{P(b)}+\frac{1}{P(c)}} \cdot \frac{\frac{1}{P(b)}}{\frac{1}{P(c)}+\frac{1}{P(b)}} \tag{3.2}$$

Computing the ratio between Equations (3.1) and (3.2), we obtain:

$$\frac{P(a)}{P(c)} \cdot \frac{P(a)+P(b)}{P(c)+P(b)} = \frac{\mathbb{P}[a \succ b \succ c]}{\mathbb{P}[c \succ b \succ a]} = \frac{P(a)^2}{P(c)^2} \cdot \frac{P(c)+P(b)}{P(a)+P(b)} \tag{3.3}$$

Defining $f : x \mapsto \frac{(1+x)^2}{x}$, Equation (3.3) can be rewritten into $f(\frac{P(a)}{P(b)}) = f(\frac{P(c)}{P(b)})$. Assuming without loss of generality that $P(b) \leq P(a), P(c)$, and because f is increasing on $[1, +\infty)$, we deduce that $P(a) = P(c)$. Symmetrically, one can show that that $f(\frac{P(b)}{P(a)}) = f(\frac{P(c)}{P(a)})$, and thus $P(a) = P(b)$, which proves that $\succ$ is uniform. □

3.2 Utility preferences

A more general class of random preferences are order statistics models. This class has been first considered by Thurstone [Thu27] and later studied by Daniels [Dan50]. The main postulate is that the person establishing a ranking observes a stimuli X_i for each option i, and sorts them by increasing order of X_i. If we allow arbitrary dependencies between X_i any distribution over rankings can be obtained, thus the assumption is made that X_i's are independent. We will call those models *utility preferences*, as they are well suited to model cardinal utilities, first considered by Pareto [Par19].

Definition 3.5 (Utility preferences)**.** Someone has utility preferences if they draw independently a (continuous) random utility U_a for each agent $a \in A$ they find acceptable, then sort acceptable partners by decreasing utility.

Definition 3.6 (Additive utility preferences)**.** Utility preferences are additive if utilities are equal to constants plus identical shocks, that is if all $U_a - \mathbb{E}[U_a]$ are identically distributed.

Interestingly, we now show that popularity and anti-popularity preferences are additive utility preferences. This property is discussed in [Luc77], and will be useful to us to give a general definition of aligned and symmetric preferences.

Lemma 3.7. *Popularity and anti-popularity preferences are additive utility preferences.*

Proof. Popularity preferences are also known as Luce model [Luc59; Luc77]. Given a popularity function $P : A \to \mathbb{R}_+$, we define an exponential random variable $X_a \sim \text{Exp}(P(a))$ for each agent $a \in A$ with positive popularity. We will show that sorting agents by increasing X_a yields the same distribution as popularity preferences.

- We define $X = \min_{a \in A} X_a$ and $Y = \arg\min_{a \in A} X_a$. Classical results on exponential random variable show that X and Y are independent, that $X \sim \text{Exp}(\sum_{a \in A} P(a))$, and that $\mathbb{P}[Y = y] = P(y)/\sum_{a \in A} P(a)$ for all $y \in A$.

- Moreover, the memory-less property of exponential variables says that X_a has the same distribution as $(X_a - x)$ when conditioning on the fact that $X_a > x$ with $x \in \mathbb{R}_+$.
- We first draw (X, Y), let the best agent be Y, and condition on the fact that $X_a > X$ for every other $a \neq Y$. We proceed by induction, and show that we obtain exactly the same distribution as popularity preferences.

To obtain additive utility preferences, we set $U_a = -\ln(X_a)$, and we have

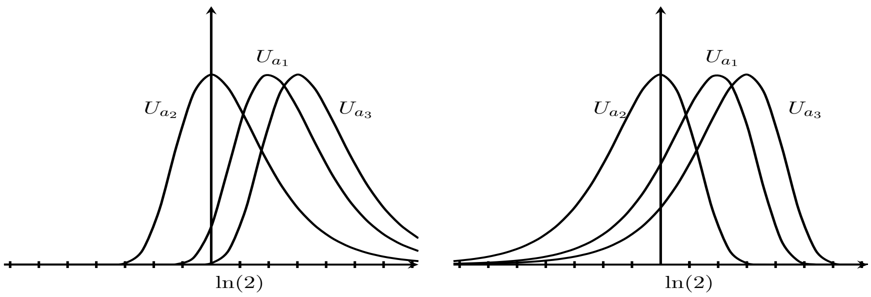
$$\forall t \in \mathbb{R}, \qquad \mathbb{P}[U_a \leq t] = \mathbb{P}[X_a \geq e^{-t}] = e^{-P(a)e^{-t}} = F(t - \ln(P(a))).$$

where $F : x \mapsto e^{-e^{-x}}$ is the cumulative distribution function of a Gumbel distribution (also known as double-exponential distribution), whose probability density function is $f : x \mapsto e^{-(x+e^{-x})}$, and whose expected value is the Euler-Mascheroni constant $\gamma \approx 0.5772$. Thus $U_a - \ln(P(a))$ are Gumbel random variables, and $U_a - \mathbb{E}[U_a]$ are identically distributed.

Symmetrically, anti-popularity preferences can be obtained by sorting agents by decreasing $X_a \sim \text{Exp}(1/P(a))$. To obtain additive utility preferences we set $U_a = \ln(X_a)$, and we have

$$\forall t \in \mathbb{R}, \qquad \mathbb{P}[U_a \leq t] = \mathbb{P}[X_a \geq e^{t}] = e^{-e^{t}/P(a)} = F(\ln(P(a)) - t).$$

Thus $\ln(P(a)) - U_a$ are Gumbel random variables, and $U_a - \mathbb{E}[U_a]$ are identically distributed. □



Popularity preferences   Anti-popularity preferences

Figure 3.5. Popularity and anti-popularity preferences are additive utility preferences. Each panel plots the probability density functions of U_{a_1}, U_{a_2} and U_{a_3}, when three agents $\{a_1, a_2, a_3\}$ have respectively popularity $P(a_1) = 4$, $P(a_2) = 1$ and $P(a_3) = 8$.

We now define the conditional-monotonicity, which is a rather natural regularity condition: if we know that a_2 is ranked first among $a_2, \ldots, a_k$, then it is less likely that a person a_1 will be ranked before a_2. This property will be used in Chapter 4.

Definition 3.8 (Conditionally-monotone distribution)**.** A distribution of preferences lists is conditionally-monotone when for every sequence of acceptable partners $a_1, \ldots, a_k$ we have $\mathbb{P}[a_1 \succ a_2 \,|\, a_2 \succ \cdots \succ a_k] \leq \mathbb{P}[a_1 \succ a_2]$.

Lemma 3.9. *Utility preferences are conditionally-monotone*

Proof. Given a random utility U and two real constants $x \geq y$, we have

$$\mathbb{P}[x > U > y] = \mathbb{P}[x > U] + \mathbb{P}[U > y] - 1 \leq \mathbb{P}[x > U] \cdot \mathbb{P}[U > y].$$

Observe that the inequality also holds if $x < y$. For each i, let us denote U_i the utility given to agent a_i and let x_i be a constant. The equation above can be rewritten to

$$\mathbb{P}[x_1 > U_2 > x_3 > \cdots > x_k] \leq \mathbb{P}[x_1 > U_2] \cdot \mathbb{P}[U_2 > x_3 > \cdots > x_k].$$

We conclude the proof, integrating over $x_1, x_3, \ldots, x_k$ using the law of total probability, and applying Fubini's Theorem to the right-hand-side of the inequality

$$\mathbb{P}[U_1 > U_2 > U_3 > \cdots > U_k] \leq \mathbb{P}[U_1 > U_2] \cdot \mathbb{P}[U_2 > U_3 > \cdots > U_k].$$

□

3.3 Correlated preferences

To model "one sided" correlations, we define aligned preferences. As an example, consider the market of PhDs and post-doc positions. All the PhDs might prefer university X to university Y, and all the universities might prefer Alice to Bob.

Definition 3.10 (Aligned preferences)**.** A set of agents have aligned utility preferences if the utilities they give to the same acceptable partner are identically distributed.

In order to model "cross sided" correlations, we introduce symmetric preferences. Imagine that Alice has a very good thesis in Computer Science, because of her skills she will most likely apply for a post-doc in a very good computer science department; symmetrically this university will most likely rank Alice first.

Definition 3.11 (Symmetric preferences)**.** Agents have symmetric utility preferences if for every pair of agents from opposite sides who find each other acceptable, the utilities they give to each other are identically distributed.

In Lemma 3.12, we show that when all agents have additive utility preferences, such that idiosyncratic shocks are identically distributed, then symmetric preferences are strictly more general than aligned preferences.

Lemma 3.12. *Assume that agents have additive utility preferences with identical shocks. If both sides have aligned preferences, then they have symmetric preferences.*

Proof. Assume that $U - \mathbb{E}[U]$ has distribution $\nu \in \Delta(\mathbb{R})$ for each utility U. Using the fact that preferences are aligned, define u_a the expected utility given to agent a. We set the utility an agent a will give to an acceptable partner b to $U_{a,b} = u_a + u_b + X_{a,b}$ with $X_{a,b} \sim \nu$. Two agents from opposite side will give each other identically distributed utilities, and the resulting preferences are identical to the original aligned preferences. □

Part II

Who can manipulate?

4 | Strongly Correlated Preferences

This chapter is based on the following paper:

[GMM21b] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Two-Sided Matching Markets with Strongly Correlated Preferences". In: *Fundamentals of Computation Theory*. Springer. 2021, pp. 3–17

4.1 Introduction

As discussed in Chapter 2 there exists instances of the stable matching problem where the men-optimal and women-optimal stable matchings are different. This raises the question of which matching to choose [Gus87; GI89] and of possible strategic behavior [DF81; Rot82; DGS87]. More precisely, we showed in Section 2.7 that if a woman lies about her preference list, this gives rise to new stable matchings, where she will be no better off than she would be in the true women-optimal matching. Thus, a woman can only gain from strategic manipulation up to the maximum difference between her best and worst partners in stable matchings.

Fortunately, there is empirical evidence that in many instances, in practice the stable matching is essentially unique (a phenomenon often referred to as "core-convergence"); see for example [RP99; PS08; HHA10; Ban+13]. One of the empirical explanations for core-convergence given by Roth and Peranson in [RP99] is that the preference lists are correlated: *"One factor that strongly influences the size of the set of stable matchings is the correlation of preferences among programs and among applicants. When preferences are highly correlated (i.e., when similar programs tend to agree which are the most desirable applicants, and applicants tend to agree which are the most desirable programs), the set of stable matchings is small."*

Following that direction of enquiry, we study the core-convergence phenomenon, in a model where preferences are stochastic. When preferences of women are strongly correlated (vertical), Theorem 4.1 shows that the expected difference of rank between each woman's worst and best stable partner is a constant, hence the incentives to manipulate are limited. If additionally the preferences of men are uncorrelated (horizontal), Theorem 4.15 shows that most women have a unique stable partner, and therefore have no incentives to manipulate. To complete the scenery, Theorem 4.19 shows that when both sides have strongly correlated preferences (vertical), stable matchings are assortative.

Related work. In a closely related paper, Holzman and Samet [HS14] look at a deterministic setting. Reindexing a summation, they show that if each man m is given almost the same rank by all women (rank difference $\leq \delta$), then women give in average almost the same ranks to their worst and best stable partners (average rank difference $\leq \delta$). This result is closely related to Theorem 4.1, where we bound individually the difference of rank of each woman's worst and best stable partners. Additionally, if each person (men and women) is given almost the same rank by others (rank difference $\leq \delta$), then two persons paired in a stable matching give each other similar ranks (rank difference $\leq 2\delta$). This is closely related to Theorem 4.19, where we bound the difference of index $|i-j|$ between a man m_i and a woman w_j matched in a stable matching. Using such techniques in a stochastic setting would give high probability bounds which depends on the number of agents. Instead, we are able to give constant upper bounds on expected values.

Analyzing instances that are less far-fetched than in the worst case is the motivation underlying the model of stochastically generated preference lists. A series of papers [Pit89; KMP90; Pit92; PSV07; LP09] study the model where N men and N women have complete uniformly random preferences. Asymptotically, and in expectation, a fixed woman w gives rank $\sim \ln N$ to her best stable husband, and rank $\sim N/\ln N$ to her worst stable husband. More recent papers have studied the robustness of those results to variations around the uniform model. Ashlagi, Kanoria and Leshno [AKL17], Kanoria, Min and Qian [KMQ21], and Ashlagi, Braverman, Saberi, Thomas and Zhao [Ash+21] study the rank of each person's partner, under the men and women optimal stable matchings, as a function of the market imbalance [AKL17], the size of preference lists [KMQ21], or as a function of each person's (bounded) popularity [Ash+21]. Even if the techniques involved are quite different, such results can be compared to Theorem 4.1, which bounds the difference of rank between a woman's worst and best stable partner.

The first theoretical explanations of the "core-convergence" phenomenon where given in [IM15] and [AKL17]. Immorlica and Mahdian [IM15] consider the case where men have constant size random preferences (truncated popularity preferences). Ashlagi, Kanoria and Leshno [AKL17], consider slightly unbalanced matching markets ($M < W$). Both articles prove that the fraction of persons with several stable partners tends to 0 as the market grows large. Theorem 4.15 and its proof incorporate ideas from those two papers.

Beyond strong "core-convergence", where most agents have a unique stable partner, one can try to compute what each person gets. Lee [Lee16] considers a model with random cardinal utilities: when an agent is matched with a partner, its utility is a function of the partner intrinsic value (which induce vertical preferences) and of a private idiosyncratic value (which induces horizontal preferences). Lee shows that stable matching are assortative and match agents having similar public values, which effectively bounds the difference in utility between each person worst and best stable partner. Such results can be compared with Theorem 4.19, and serve as a transition to the second part of this thesis which is dedicated to the question "who gets what?".

Beyond one-to-one matchings, school choice is an example of many-to-one markets. Kojima and Pathak [KP09] generalize results from [IM15] and prove that most schools have no incentives to manipulate. Azevedo and Leshno [AL16] show that large markets converge to a unique stable matching in a model with a continuum of students. To counter balance those findings, Biró, Hassidim, Romm and Shorer [Bir+20], and Rheingans-Yoo [RS20] argue that socioeconomic status and geographic preferences might undermine core-convergence, thus some incentives remain in such markets.

Takeaway message. It is well established that incentives of agents are related to how "balanced" the market is. When the market is unbalanced, some agents from the large side will be single, and agents from the small side will never be matched to agents to whom they prefer someone single. Therefore, only stable matchings which are nearly optimal for the small side will remain. Ashlagi, Kanoria and Leshno [AKL17] showed that removing one man is enough for the set of stable matching to collapse. Immorlica and Mahdian [IM15] observed this phenomenon when agents have small preference lists. Conversely, Biró, Hassidim, Romm and Shorer [Bir+20], and Rheingans-Yoo [RS20] show that "locally balanced" structures create an opportunity for multiple stable matchings. We argue that this fact gives an intuitive interpretation of our results. When one side of the market has vertical (strongly correlated) preferences and the other side has horizontal (uniform) preferences, the market is highly unbalanced, which explains the strong conclusion of Theorem 4.15. When both sides have vertical preferences we observe locally balanced groups, which explain the relatively weaker conclusion of Theorem 4.19.

4.2 Vertical - Arbitrary preferences

When women have strongly correlated preferences, it induces a canonical ordering of men. Informally speaking, we will refer to this situation as "vertical preferences". The following theorem shows that every woman gives approximately the same rank to all of her stable partners.

Theorem 4.1. *Assume that each woman independently draws her preference list from a conditionally-monotone distribution. The men's preference lists are arbitrary. Let u_k be an upper bound on the odds that man m_{i+k} is ranked before man m_i:*

$$\forall k \geq 1, \quad u_k = \max_{w,i} \left\{ \frac{\mathbb{P}[m_{i+k} \succ_w m_i]}{\mathbb{P}[m_i \succ_w m_{i+k}]} \;\middle|\; w \text{ finds both } m_i \text{ and } m_{i+k} \text{ acceptable} \right\}$$

Then for each woman with at least one stable partner, in expectation all of her stable partners are ranked within $(1 + 2\exp(\sum_{k\geq 1} k u_k)) \sum_{k \geq 1} k^2 u_k$ of one another in her preference list.

Theorem 4.1 is most relevant when the women's preference lists are strongly correlated, that is, when every woman's preference list is "close" to a single ranking $m_1 \succ m_2 \succ \ldots \succ m_M$. This closeness is measured by the odds that in some ranking, some man is ranked ahead of a man who, in the ranking $m_1 \succ m_2 \succ \ldots \succ m_M$, would be k slots ahead of him.

We detail below three examples of applications, where the expected difference of ranks between each woman's best and worst partners is $O(1)$, and thus her incentives to misreport her preferences are limited.

- *Identical preferences.* If all women rank their acceptable partners using a master list $m_1 \succ m_2 \succ \cdots \succ m_M$, then all u_k's are equal to 0. Then Theorem 4.1 states that each woman has a unique stable husband, a well-known result for this type of instances.

- *Preferences from identical popularities.* Assume that women have popularity preferences (Definition 3.2) and that each woman gives man m_i popularity 2^{-i}. Then $u_k = 2^{-k}$ and the expected rank difference is at most $\mathcal{O}(1)$.

- *Preferences from correlated utilities.* Assume that women have additive utility preferences (Definition 3.6), where each woman w gives man m_i a score that is the sum of a common value i and an idiosyncratic value η_i^w which is normally distributed with mean 0 and variance σ^2; she then sorts men by increasing scores. Then we have the upper-bound $u_k \leq \max_{w,i} \{2 \cdot \mathbb{P}[\eta_i^w - \eta_{i+k}^w > k]\} \leq 2e^{-(k/2\sigma)^2}$ and the expected rank difference, by a short calculation, is at most $4\sqrt{\pi}\sigma^3(1 + 2e^{4\sigma^2}) = \mathcal{O}(1)$.

In Section 4.2.1, we define a partition of stable matching instances into *tiers*. For strongly correlated instances, tiers provide the structural insight to start the analysis: in Lemma 4.6, we use them to upper-bound the difference of ranks between a woman's worst and best stable partners by the sum of (1) the number x of men coming from other tiers and who are placed between stable husbands in the woman's preference list, and (2) the tier size.

The analysis requires a delicate handling of conditional probabilities. In Section 4.2.2, we explain how to condition on the men-optimal stable matching, when preferences are random.

Section 4.2.3 analyzes (1). The men involved are out of place compared to their position in the ranking $m_1 \succ \ldots \succ m_M$, and the odds of such events can be bounded, thanks to the assumption that distributions of preferences are conditionally-monotone. Our main technical lemma there is Lemma 4.8.

Section 4.2.4 analyzes (2), the tier size by first giving a simple greedy algorithm (Algorithm 4.1) to compute a tier. Each of the two limits of a tier is computed by a sequence of "jumps", so the total distance traveled is a sum of jumps which, thanks to Lemma 4.8 again, can be stochastically dominated by a sum X of independent random variables (see Lemma 4.12); thus it all reduces to analyzing X, a simple mathematical exercise (Lemma 4.13).

Finally, Section 4.2.5 combines the Lemmas previously established to prove Theorem 4.1. Our analysis builds on Theorems 2.3 and 2.4, two fundamental and well-known results.

4.2.1 Separators and tiers

In this subsection, we define the tier structure underlying our analysis.

Definition 4.2 (separator)**.** A *separator* is a set $S \subseteq \mathcal{M}$ of men such that in the men-optimal stable matching $\mu_{\mathcal{M}}$, each woman married to a man in S prefers him to all men outside S :

$$\forall w \in \mu_{\mathcal{M}}(S) \cap \mathcal{W}, \quad \forall m \in \mathcal{M} \setminus S, \quad \mu_{\mathcal{M}}(w) \succ_w m$$

Lemma 4.3. *Given a separator $S \subseteq \mathcal{M}$, every stable matching matches S to the same set of women.*

Proof. Let $w \in \mu_{\mathcal{M}}(S)$ and let m be the partner of w in some stable matching. Since $\mu_{\mathcal{M}}$ is the woman-pessimal stable matching by Theorem 2.3, w prefers m to $\mu_{\mathcal{M}}(w)$. By definition of separators, that implies that $m \in S$. Hence, in every stable matching μ, women of $\mu_{\mathcal{M}}(S)$ are matched to men in S. By a cardinality argument, men of S are matched by μ to $\mu_{\mathcal{M}}(S)$. □

Definition 4.4 (prefix separator, tier)**.** A *prefix separator* is a separator S such that $S = \{m_1, m_2, \ldots, m_t\}$ for some $0 \leq t \leq N$. Given a collection of $b+1$ prefix separators $S_i = \{m_1, \ldots, m_{t_i}\}$ with $0 = t_0 < t_1 < \cdots < t_b = N$, the i-th *tier* is the set $B_i = S_{t_i} \setminus S_{t_{i-1}}$ with $1 \leq i \leq b$. Abusing notations, we will denote S as the prefix separator t and B as the tier $(t_{i-1}, t_i]$. See Figure 4.1 for an illustration.

Lemma 4.5. *Given a tier $B \subseteq \mathcal{M}$, every stable matching matches B to the same set of women.*

Proof. B equals $S_{t_i} \setminus S_{t_{i-1}}$ for some i. We apply Lemma 4.3 to S_{t_i} and to $S_{t_{i-1}}$. □

Lemma 4.6. *Consider a woman w who is matched with man m_n by $\mu_{\mathcal{M}}$ and let $B = (l, r]$ denote her tier. Let x denote the number of men from a better tier that are ranked by w between a man of B and m_n:*

$$x = |\{i \leq l \mid \exists j > l,\ m_j \succ_w m_i \succ_w m_n\}|.$$

Then in w's preference list, the difference of ranks between w's worst and best stable partners is at most $x + r - l - 1$.

Proof. Since $\mu_{\mathcal{M}}$ is woman-pessimal by Theorem 2.3, m_n is the last stable husband in w's preference list. Let m_j denote her best stable husband.

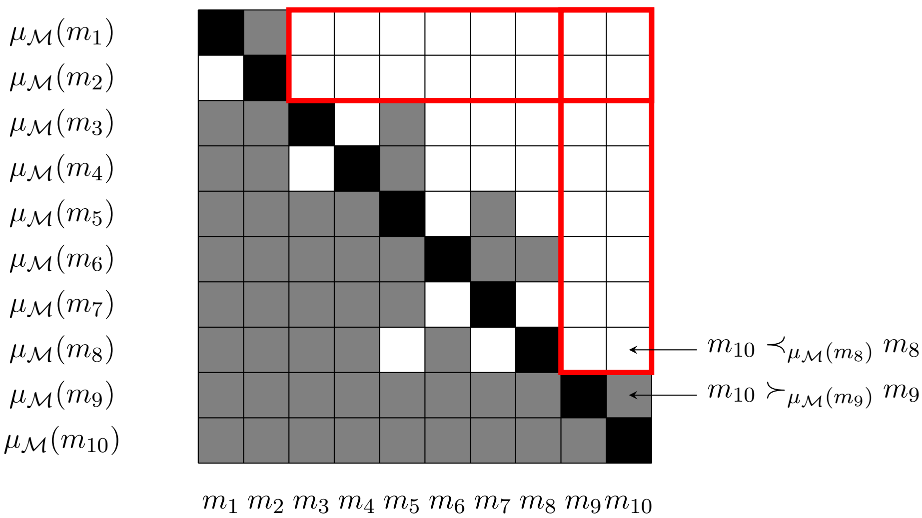


Figure 4.1. Graphical representation of prefix separators. Black cells corresponds to the men optimal stable matching $\mu_{\mathcal{M}}$. Each gray cell corresponds to "half of a blocking pair", where a woman prefer a man to her husband in $\mu_{\mathcal{M}}$. Prefix separators define 3 tiers: $(0, 2]$, $(2, 8]$ and $(8, 10]$

.

In w's preference list, the interval from m_j to m_n contains men from her own tier, plus possibly some additional men. Such a man m_i comes from outside her tier $(l, r]$ and she prefers him to m_n: since r is a prefix separator, we must have $i \leq l$. Thus x counts the number of men who do not belong to her tier but who in her preference list are ranked between m_j and m_n.

On the other hand, the number of men who belong to her tier and who in her preference list are ranked between m_j and m_n (inclusive) is at most $r - l$.

Together, the difference of ranks between w's worst and best stable partners is at most $x + (r - l) - 1$. See Figure 4.2 for an illustration. □

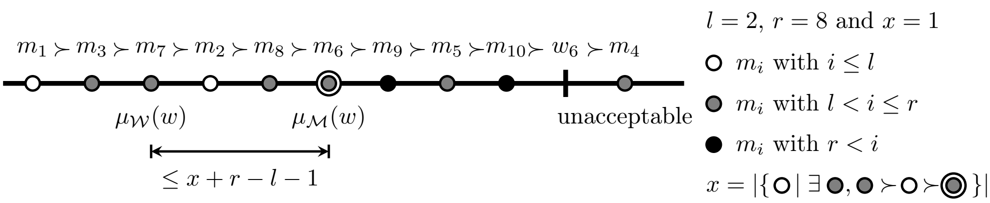


Figure 4.2. Preference list of $w = \mu_{\mathcal{M}}(m_n)$ with $n = 6$. The tier of w is defined by a left separator at $l = 2$ and a right separator at $r = 8$. Colors white, gray and black corresponds to tiers, and are defined in the legend. All stable partners of w must be gray. Men in black are all ranked after $m_n = \mu_{\mathcal{M}}(w)$. The difference in rank between w's worst and best partner is at most the number of gray men (here $r - l = 6$), minus 1, plus the number of white men ranked after a gray man and before m_n (here $x = 1$).

4.2.2 Conditioning when preferences are random

We study the case where each person draws her preference list from an arbitrary distribution. The preference lists are random variables, that are independent but not necessarily identically distributed.

Intuitively, we use the *principle of deferred decision* and construct preference lists in an online manner. By Theorem 2.3 the man-optimal stable matching $\mu_{\mathcal{M}}$ is computed by Algorithm 2.1, and the remaining randomness can be used for a stochastic analysis of each person's stable partners. To be more formal, we define a random variable $\mathcal{H}$, and inspection of Algorithm 2.1 shows that $\mathcal{H}$ contains enough information on each person's preferences to run Algorithm 2.1 deterministically.

Definition 4.7. Let $\mathcal{H} = (\mu_{\mathcal{M}}, (\sigma_m)_{m\in\mathcal{M}}, (\pi_w)_{w\in\mathcal{W}})$ denote the random variable consisting of (1) the man-optimal stable matching $\mu_{\mathcal{M}}$, (2) each man's ranking of the women he prefers to his partner in $\mu_{\mathcal{M}}$, and (3) each woman's ranking of the men who prefer her to their partner in $\mu_{\mathcal{M}}$.

4.2.3 Analyzing the number x of men from other tiers

Lemma 4.8. *Recall the sequence $(u_k)_{k\geq 1}$ defined in the statement of Theorem 4.1:*

$$\forall k \geq 1, \quad u_k = \max_{w,i} \left\{ \frac{\mathbb{P}[m_{i+k} \succ_w m_i]}{\mathbb{P}[m_i \succ_w m_{i+k}]} \;\middle|\; w \text{ finds both } m_i \text{ and } m_{i+k} \text{ acceptable} \right\}$$

Let w be a woman. Given a subset of her acceptable men and a ranking of that subset $a_1 \succ_w \cdots \succ_w a_p$, we condition on the event that in w's preference list, $a_1 \succ_w \cdots \succ_w a_p$ holds. Let $m_i = a_1$ be w's favorite man in that subset. Let J_i be a random variable, equal to the highest $j \geq i$ such that woman w prefers m_j to m_i. Formally, $J_i = \max\{j \geq i \mid m_j \succeq_w m_i\}$. Then, for all $k \geq 1$, we have

$$\mathbb{P}[J_i < i+k \mid J_i < i+k+1] \geq \exp(-u_k), \quad \textit{and} \quad \mathbb{P}[J_i < i+k] \geq \exp(-\textstyle\sum_{\ell\geq k} u_\ell).$$

Proof. J_i is determined by w's preference list. We construct w's preference list using the following algorithm: initially we know her ranking σ_A of the subset $A = \{a_1, a_2, \ldots, a_p\}$ of acceptable men, and $m_i = a_1$ is her favorite among those. For each j from N to i in decreasing order, we insert m_j into the ranking according to the distribution of w's preference list, stopping as soon as some m_j is ranked before m_i (or when $j = i$ is that does not happen). Then the step $j \geq i$ at which this algorithm stops equals J_i.

To analyze the algorithm, observe that at each step $j = N, N-1, \ldots$, we already know w's ranking of the subset $S = \{m_{j+1}, \ldots, m_N\} \cup \{a_1, \ldots, a_p\} \cup \{\text{men who are not acceptable to } w\}$. If m_j is already in S, w prefers m_i to m_j, thus the algorithm continues and $J_i < j$. Otherwise the algorithm inserts m_j into the existing ranking: by definition of conditionally-monotone distributions (Definition 3.8), the probability that m_j beats m_i given the ranking constructed so far is at most the unconditional probability $\mathbb{P}[m_j \succ_w m_i]$.

$$\mathbb{P}[J_i < j \mid w\text{'s partial ranking at step } j] \geq 1 - \mathbb{P}[m_j \succ_w m_i].$$

By definition of u_{j-i}, we have $1 - \mathbb{P}[m_j \succ_w m_i] = \left(1 + \frac{\mathbb{P}[m_j \succ_w m_i]}{\mathbb{P}[m_i \succ_w m_j]}\right)^{-1} \geq (1+u_{j-i})^{-1} \geq \exp(-u_{j-i})$.

Summing over all rankings σ_S of S that are compatible with σ_A and with $J_i \leq j$,

$$\mathbb{P}[J_i < j \mid J_i \leq j] = \sum_{\substack{\sigma_S \text{ compatible with} \\ J_i \leq j \text{ and with } \sigma_A}} \mathbb{P}[\sigma_S \mid \sigma_A] \cdot \mathbb{P}[J_i < j \mid \sigma_S]$$

$$\geq \sum_{\sigma_S} \mathbb{P}[\sigma_S \mid \sigma_A] \cdot \exp(-u_{j-i}) = \exp(-u_{j-i}).$$

Finally, $\mathbb{P}[J_i < j] = \prod_{\ell=j}^{N} \mathbb{P}[J_i < \ell \mid J_i \leq \ell] \geq \prod_{k \geq j-i} \exp(-u_k)$. □

Recall from Lemma 4.6 that $r - l - 1 + x$ is an upper bound on the difference of rank of woman w's worst and best stable husbands. We first bound the expected value of the random variable x defined in Lemma 4.6.

Lemma 4.9. *Given a woman $w = \mu_{\mathcal{M}}(m_n)$, define the random variable x as in Lemma 4.6: conditioning on $\mathcal{H}$, $x = |\{i \leq l \mid \exists j > l,\ m_j \succ_w m_i \succ_w m_n\}|$ is the number of men in a better tier, who can be ranked between w's worst and best stable husbands. Then $\mathbb{E}[x] \leq \sum_{k \geq 1} k u_k$.*

Proof. Start by conditioning on $\mathcal{H}$, and let $m_n = a_1 \succ_w a_2 \succ_w \cdots \succ_w a_p$ be w's ranking of men who prefer her to their partner in $\mu_{\mathcal{M}}$. We draw the preference lists of each woman w_i with $i < n$, and use Algorithm 4.1 to compute the value of l.

For each $i \leq l$, we proceed as follows. If $m_n \succ_w m_i$, then m_i cannot be ranked between w's worst and best stable partners. Otherwise, we are in a situation where $m_i \succ_w a_1 \succ w \cdots \succ_w a_p$. Using notations from Lemma 4.8, w prefers m_i to all m_j with $j > l$ if and only if $J_i < l + 1$. By Lemma 4.8 this occurs with probability at least $\exp(-\sum_{k \geq l+1-i} u_k)$. Thus

$$\mathbb{P}[\exists j > l,\ m_j \succ_w m_i \succ_w m_n \mid \mathcal{H}, l] \leq 1 - \exp(-\textstyle\sum_{k \geq l+1-i} u_k) \leq \sum_{k \geq l+1-i} u_k.$$

Summing this probability for all $i \leq l$, we obtain $\mathbb{E}[x \mid \mathcal{H}, l] \leq \sum_{i \leq l} \sum_{k \geq l+1-i} u_k \leq \sum_{k \geq 1} k u_k$. □

4.2.4 Analyzing the tier size

Lemma 4.10. *Given n such that man m_n is matched in in $\mu_{\mathcal{M}}$, Algorithm 4.1 outputs the tier $(l, r]$ containing n.*

Algorithm 4.1 Computing a tier

Initialization:
Compute the man optimal stable matching $\mu_{\mathcal{M}}$.
Relabel women so that w_i denotes the wife of m_i in $\mu_{\mathcal{M}}$
Let n be such that man $w_n = \mu_{\mathcal{M}}(m_n)$.
Left prefix separator: initialize $l \leftarrow n - 1$
While there exists $i \leq l$ and $j > l$ such that $m_j \succ_{w_i} m_i$, **do**
  $l \leftarrow \min\{i \leq l \mid \exists j > l,\ m_j \succ_{w_i} m_i\} - 1$.
Right prefix separator: initialize $r \leftarrow n$.
While there exists $j > r$ and $i \leq r$ such that $m_j \succ_{w_i} m_i$, **do**
  $r \leftarrow \max\{j > r \mid \exists i \leq r,\ m_j \succ_{w_i} m_i\}$.
Output: $(l, r]$.

Proof. Algorithm 4.1 is understood most easily by following its execution on Figure 4.3. Algorithm 4.1 applies a right-to-left greedy method to find the largest prefix separator l which is $\leq n-1$. By definition of prefix separators, a witness that some t is not a prefix separator is a pair (m_j, w_i) where $j > t \geq i$ and woman w_i prefers man m_j to her partner: $m_j >_{w_i} m_i$. Then the same pair also certifies that no $t' = t, t-1, t-2, \ldots, i$ can be a prefix separator either, so the algorithm jumps to $i-1$ and looks for a witness again. When there is no witness, a prefix separator has been found, thus l is the largest prefix separator $\leq n-1$. Similarly, Algorithm 4.1 computes the smallest prefix separator r which is $\geq n$. Thus, by definition of tiers, $(l, r]$ is the tier containing w_n. □

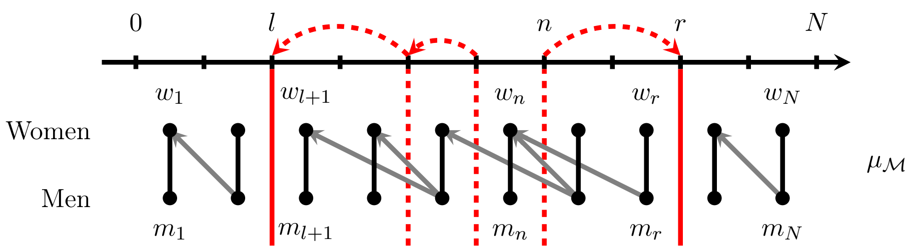


Figure 4.3. Computing the tier containing n. The vertical black edges correspond to the men-optimal stable matching $\mu_{\mathcal{M}}$. There is a light gray arc (m_j, w_i) if $j > i$ and woman w_i prefers man m_j to her partner: $m_j \succ_{w_i} m_i$. The prefix separators correspond to the solid red vertical lines which do not intersect any gray arc. Algorithm 4.1 applies a right-to-left greedy method to find the largest prefix separator l which is $\leq n-1$, jumping from dashed red line to dashed red line, and a similar left-to-right greedy method again to find the smallest prefix separator r which is $\geq n$. This determines the tier $(l, r]$ containing n.

Definition 4.11. Let X be the random variable defined as follows. Let $(\Delta_t)_{t\geq 0}$ denote a sequence of i.i.d.r.v.'s taking non-negative integer values with the following distribution:

$$\forall \delta > 0, \quad \mathbb{P}[\Delta_t < \delta] = \exp\left(-\sum_{k\geq\delta} k u_k\right)$$

Then $X = \Delta_0 + \Delta_1 + \cdots + \Delta_{T-1}$, where T is the first $t \geq 0$ such that $\Delta_t = 0$.

Lemma 4.12. *Given a woman $w = \mu_{\mathcal{M}}(m_n)$, let $(l, r]$ denote the tier containing n. Conditioning on $\mathcal{H}$, l and r are integer random variable, such that $r-n$ and $n-1-l$ are stochastically dominated by X.*

Proof. Conditioning on $\mathcal{H}$, we know the men-optimal stable matching $\mu_{\mathcal{M}}$, and each woman's ranking of the men who prefer her to their partner in $\mu_{\mathcal{M}}$. We relabel women so that w_i denotes the wife of m_i in $\mu_{\mathcal{M}}$. Using notations from Lemma 4.8, let $J_i = \max\{j \geq i \mid m_j \succeq_{w_i} m_i\}$, for all $1 \leq i \leq M$.

We start with a stochastic domination of $r-n$. From Lemma 4.10, the right separator r is computed with a while loop. Let $r_0 = n$ be the initial value of r. To decide whether r_0 is a separator, we look at w_n's preference list. Let $r_1 = J_n$ be the maximum $j \geq n$ such that w_n prefer m_j to m_n. If $r_1 = r_0$, w_n prefers m_n to all men m_j with $j > n$, and r_0 is a prefix separator. Otherwise, no prefix separator can exist between r_0 and r_1. Using Lemma 4.8, $r_1 - r_0$ is stochastically dominated by Δ_0.

$$\forall \delta > 0, \quad \mathbb{P}[r_1 - r_0 < \delta \mid \mathcal{H}] = \mathbb{P}[J_n < n + \delta \mid \mathcal{H}] \geq \exp(-\sum_{k\geq\delta} u_k) \geq \mathbb{P}[\Delta_0 < \delta]$$

For all $t > 0$, we proceed by induction. To decide whether r_t is a separator, we look at the preference lists of $w_{1+r_{t-1}}, \dots, w_{r_t}$. Let $r_{t+1} = \max\{J_{1+r_{t-1}}, \dots, J_{r_t}\}$ be the maximum $j \geq r_t$ such that a woman w_i prefer m_j to m_i, with $r_{t-1} < i \leq r_t$. If $r_{t+1} = r_t$, then r_t is a prefix separator. Otherwise, no prefix separator can exist between r_t and r_{t+1}. We show that Δ_t stochastically dominates $r_{t+1} - r_t$.

$$\begin{aligned}
&\forall \delta > 0, \mathbb{P}[r_{t+1} - r_t < \delta \mid \mathcal{H}, J_n, \dots, J_{r_{t-1}}] \\
&= \prod_{i=1+r_{t-1}}^{r_t} \mathbb{P}[J_i < r_t + \delta \mid \mathcal{H}, J_n, \dots, J_{r_{t-1}}] \\
&\geq \prod_{i=1+r_{t-1}}^{r_t} \exp(-\textstyle\sum_{k \geq r_t+\delta-i} u_k) \quad \text{(Lemma 4.8)} \\
&\geq \exp(-\textstyle\sum_{k \geq \delta} k u_k) = \mathbb{P}[\Delta_t < \delta]
\end{aligned}$$

Summing up to t such that $r_{t+1} = r_t$ proves that X stochastically dominates $r - n$.

We now prove that X stochastically dominates $n - 1 - l$. From Lemma 4.10, the left separator l is computed with a while loop, and let $l_0 = n - 1$ be its initial value. To decide whether l_0 is a prefix separator, we need to know if a woman w_i prefers a man m_j to her husband m_i, with $i \leq l_0 < j$. More formally, l_0 is a prefix separator if and only if $J_i \leq l_0$ for all $i \leq l_0$. Defining $l_1 = \min\{i \leq l_t + 1 \mid J_i > l_t\} - 1$, $l_1 = l_0$ if and only if l_0 is a separator. Using Lemma 4.8, $l_1 - l_0$ is stochastically dominated by Δ_0.

$$\begin{aligned}
\forall \delta > 0, \quad \mathbb{P}[l_0 - l_1 < \delta \mid \mathcal{H}] &= \mathbb{P}[J_1, \dots, J_{l_0-\delta+1} l_0 \mid \mathcal{H}] \\
&\geq \exp(-\textstyle\sum_{k \geq \delta} u_k) \\
&\geq \mathbb{P}[\Delta_0 < \delta]
\end{aligned}$$

For all $t > 0$, we proceed by induction and let $l_{t+1} = \min\{i \leq l_t + 1 \mid J_i > l_t\} - 1$. More precisely, $l_{t+1} + 1$ is the minimum $i \leq l_t + 1$ such that w_i prefer a man m_j to her husband m_i with $j > l_t$. If $l_{t+1} = l_t$, then l_t is a prefix separator, and the process stop here. Otherwise, no prefix separator can exist between l_{t+1} and l_t. A crucial property is that for all $i \leq l_{t+1}$, the best man in w_i's partial list is still m_i, hence Lemma 4.8 will still be applicable the next step.

$$\begin{aligned}
\forall \delta > 0, \quad \mathbb{P}\left[l_t - l_{t+1} < \delta \;\middle|\; \begin{matrix} J_1, \dots, J_{l_t} \leq l_{t-1} \\ \mathcal{H}, l_0, \dots, l_t \end{matrix} \right] &= \prod_{i=1}^{l_t-\delta+1} P[J_i \leq l_t \mid J_i \leq l_{t-1}, \mathcal{H}] \\
&\geq \prod_{i=1}^{l_t-\delta+1} \exp(-\textstyle\sum_{k \geq l_t+1-i} u_k) \quad \text{(Lemma 4.8)} \\
&\geq \exp(-\textstyle\sum_{k \geq \delta} k u_k) = \mathbb{P}[\Delta_t < \delta]
\end{aligned}$$

Summing up to t such that $l_{t+1} = l_t$ proves that X stochastically dominates $n - 1 - l$. □

Lemma 4.13. *We have* $\mathbb{E}[X] \leq \exp(\sum_{k \geq 1} k u_k) \sum_{k \geq 1} k^2 u_k$.

Proof. From Wald's equation, $\mathbb{E}[X] = \mathbb{E}[T] \cdot \mathbb{E}[\Delta_0]$. The random variable T is geometrically distributed, with a success parameter $\mathbb{P}[\Delta_0 = 0]$, hence $E[T] = 1/\mathbb{P}[\Delta_0 = 0]$. Because Δ_0 only takes non-negative integer values, we can compute its expectation with a sum.

$$\mathbb{E}[\Delta_0] = \sum_{\delta \geq 0} \mathbb{P}[\Delta_0 > \delta] = \sum_{\delta \geq 0} 1 - \exp\left(-\textstyle\sum_{k > \delta} k u_k\right) \leq \sum_{\delta \geq 0} \sum_{k > \delta} k u_k = \sum_{k \geq 1} k^2 u_k$$

□

Lemma 4.14. *Assuming that $u_k = \exp(-\Omega(k))$, we have $\mathbb{P}[X \geq k] = \exp(-\Omega(k))$.*

Proof. Let $G_X(z) = \mathbb{E}[z^X]$ be the probability generating function of X, which is defined at least for all real z such that $|z| < 1$. In addition if $G_X(1+\varepsilon)$ is finite for some $\varepsilon > 0$, then Markov's inequality gives

$$\forall k \geq 0, \quad \mathbb{P}[X \geq k] = \mathbb{P}[(1+\varepsilon)^X \geq (1+\varepsilon)^k] \leq G_X(1+\varepsilon)\exp(-k\ln(1+\varepsilon)) = \exp(-\Omega(k)).$$

Computing G_X using Definition 4.11, and conditioning on the value of T.

$$G_X(z) = \mathbb{E}[z^X] = \sum_{t=0}^{+\infty} \mathbb{P}[T=t] \cdot \mathbb{E}\left[z^{\sum_{i=0}^{t-1}\Delta_i} \,\middle|\, \forall i \in [0, t-1],\ \Delta_i > 0\right]$$

Using the fact that all Δ_i's are *i.i.d.* we can simplify the expectation of the product.

$$G_X(z) = \cdot \sum_{t=0}^{+\infty} \mathbb{P}[T=t] \cdot \mathbb{E}\left[z^{\Delta_0} \mid \Delta_0 > 0\right]^t = G_T\left(\mathbb{E}\left[z^{\Delta_0} \mid \Delta_0 > 0\right]\right)$$

The conditional expectation can be expressed as follows.

$$G_{\Delta_0}(z) = \mathbb{E}\left[z^{\Delta_0}\right] = \mathbb{P}[\Delta_0 > 0] \cdot \mathbb{E}\left[z^{\Delta_0} \mid \Delta_0 > 0\right] + \mathbb{P}[\Delta_0 = 0]$$

$$\mathbb{E}\left[z^{\Delta_0} \mid \Delta_0 > 0\right] = \frac{G_{\Delta_0}(z) - \mathbb{P}[\Delta_0 = 0]}{\mathbb{P}[\Delta_0 > 0]}$$

Now let us compute the generating function of T.

$$G_T(z) = \mathbb{E}[z^T] = \sum_{t=0}^{+\infty} z^t \cdot \mathbb{P}[T=t] = \sum_{k=0}^{+\infty} z^t \cdot \mathbb{P}[\Delta_0 > 0]^t \cdot \mathbb{P}[\Delta_0 = 0] = \frac{\mathbb{P}[\Delta_0 = 0]}{1 - z \cdot \mathbb{P}[\Delta_0 > 0]}$$

Combining the three previous equations we obtain

$$G_X(z) = \frac{\mathbb{P}[\Delta_0 = 0]}{1 + \mathbb{P}[\Delta_0 = 0] - G_{\Delta_0}(z)}$$

Because of the assumption on women's preference distributions, we have $u_k = \exp(-\Omega(k))$. Hence,

$$\forall \delta \geq 1, \quad \mathbb{P}[\Delta_0 = \delta] = \mathbb{P}[\Delta_0 < \delta+1] - \mathbb{P}[\Delta_0 < \delta] = \exp(-\textstyle\sum_{k>\delta} k u_k)(1 - \exp(-\delta u_\delta)) \leq \delta u_\delta = \exp(-\Omega(\delta))$$

Thus, the convergence radius of G_{Δ_0} is strictly greater than 1. Because G_{Δ_0} is a probability generating function, it is continuous, strictly increasing, and $G_{\Delta_0}(1) = 1$. Therefore, there exists $\varepsilon > 0$ such that $G_{\Delta_0}(1+\varepsilon) < 1 + \mathbb{P}[\Delta_0 = 0]$, which concludes the proof. □

4.2.5 Putting everything together

Proof of Theorem 4.1. Without loss of generality, we may assume that $N = M \leq W$ and that each man is matched in the man-optimal stable matching $\mu_{\mathcal{M}}$: to see that, for each man m we add a "virtual" woman w as his least favorite acceptable partner, such that m is the only acceptable partner of w. A man is single in the original instance if and only if he is matched to a "virtual" woman in the new instance.

We start our analysis by conditioning on the random variable $\mathcal{H}$ (see Definition 4.7). Algorithm 2.1 then computes $\mu_{\mathcal{M}}$, which matches each woman to her worst stable partner. Up to *relabeling* the women, we may also assume that for all $i \leq N$ we have $w_i = \mu_{\mathcal{M}}(m_i)$.

Let w_n be a woman who is married in $\mu_{\mathcal{M}}$. From there, we use Lemma 4.6 to bound the difference of rank between her worst and best stable partner by $x + r - l - 1 = x + (r - n) + (n - l - 1)$. We bound the expected value of x using Lemma 4.9, and the expected values of both $r - n$ and $n - l - 1$ using Lemmas 4.12 and 4.13. □

4.3 Vertical - Horizontal preferences

A stronger notion of approximate incentive compatibility is near-uniqueness of a stable matching, meaning that most persons have either no or one unique stable partner, and thus have no incentive to misreport their preferences. When does that hold? One answer is given by Theorem 4.15.

Theorem 4.15. *Assume that each woman independently draws her preference list from a conditionally-monotone distribution. Let u_k be an upper bound on the odds that man m_{i+k} is ranked before man m_i:*

$$\forall k \geq 1, \quad u_k = \max_{w,i} \left\{ \frac{\mathbb{P}[m_{i+k} \succ_w m_i]}{\mathbb{P}[m_i \succ_w m_{i+k}]} \;\middle|\; w \text{ finds both } m_i \text{ and } m_{i+k} \text{ acceptable} \right\}$$

Further assume that all preferences are complete, that $u_k = \exp(-\Omega(k))$, and that men have popularity preferences where each woman's popularity is between 1 and a constant P. Then, in expectation the fraction of persons who have multiple stable partners is $\mathcal{O}(P \ln^2 N/N)$.

Notice that in the three examples of Theorem 4.1, the sequence $(u_k)_{k\geq 1}$ is exponentially decreasing. The assumptions of Theorem 4.15 are minimal in the sense that removing one would bring us back to a case where a constant fraction of woman have multiple stable partners.

- *Preference lists of women.* If we remove the assumption that u_k is exponentially decreasing, the conclusion no longer holds: consider a balanced market balanced ($M = W$) in which both men and women have complete uniformly random preferences; then most women have $\sim \ln N$ stable husbands [KMP90; Pit92]

- *Preference lists of men.* Assume that men have random preference built as follows: starting from the ordering $w_1, w_2, \ldots, w_M$, each pair (w_{2i-1}, w_{2i}) is swapped with probability $1/2$, for all i. A symmetric definition for women's preferences satisfy the hypothesis of Theorem 4.15, with $u_1 = 1$ and $u_k = 0$ for all $k \geq 2$. Then there is a $1/8$ probability that men m_{2i-1} and m_{2i} are both stable partners of women w_{2i-1} and w_{2i}, for all i, hence a constant expected fraction of persons with multiple stable partners.

- *Incomplete preferences.* Consider a market divided into groups of size 4 of the form $\{m_{2i-1}, m_{2i}, w_{2i-1}, w_{2i}\}$, where a man and a woman are mutually acceptable if they belong to the same group. Once again, with constant probability, m_{2i-1} and m_{2i} are both stable partners of women w_{2i-1} and w_{2i}.

The proof first continues the analysis of tiers started in Section 4.2.4. When $u_k = \exp(-\Omega(k))$, it can be tightened with a mathematical analysis to prove (Lemma 4.17) that with high probability, no tier size exceeds $O(\log n)$, and that in addition, in her preference list no woman switches the relative ordering of two men m_i and $m_{i+\Omega(\log n)}$. The rest of the proof assumes that those properties hold. The only remaining source of randomness comes from the preference lists of men.

The intuition is that it is hard for man m_i to have another stable partner from his tier. Because we assume that m_i has popularity preferences with bounded popularities, his list is likely to have some woman w_j with $j \gg i$ between w_i and the next person from his tier. Woman w_j likes m_i better than her own partner, because of the no-switching property, and m_i likes her better than his putative second stable partner, so they form a blocking pair preventing m_i's second stable partner. Transforming that intuition into a proof requires care because of the need to condition on several events.

4.3.1 Typical instances

Definition 4.16. Let $C = O(1)$ be a constant to be defined later. Let $\mathcal{K}$ denote the event that every tier has size at most $C \ln N$, and every woman prefers man m_i to man m_{i+k} for every i, whenever $k \geq C \ln N$.

Lemma 4.17. *One can choose $C = \mathcal{O}(1)$ such that the probability of event $\mathcal{K}$ is $\geq 1 - 1/N^2$.*

Proof. For the first case of failure, recall from Lemma 4.14 that the size of a tier has an exponential tail. Thus we can choose C such that the probability of a given tier has a size greater than $C \log N$ is at most $1/(2N^3)$. There are at most N tiers, using the union bound the probability that at least one has a size exceeding $C \log N$ is at most $1/(2N^2)$.

For the second case of failure, notice that the probability for a woman to prefer a man m_j to a another man m_i with $j > i + C \ln N \leq j$ is at most $u_{j-i} = e^{-\Omega(j-i)} = N^{-C\Omega(1)}$. Thus we can choose C such that the probability of this happening is smaller than $1/(2N^5)$. Using the union bound over all triples of woman/m_i/m_j, the probability of a failure is at most $1/(2N^2)$.

Choosing C maximal between the two values, and using the union bound over the two possible cases of failure, the probability that $\mathcal{K}$ does not holds is at most $1/N^2$. □

4.3.2 Blocking pairs

Lemma 4.18. *Fix $i \in [1, N]$. Conditioning on $\mathcal{H}$ and on $\mathcal{K}$, the probability that woman $w = \mu_{\mathcal{M}}(m_i)$ has more than one stable husband is at most $3PC \ln N/(N + C \ln N - i)$.*

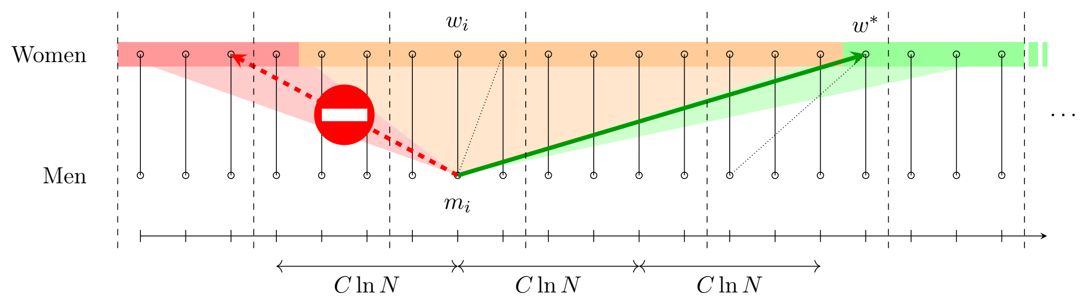


Figure 4.4. Proof of Lemma 4.18: the probability that w_i has several stable husbands is smaller than the ratio $|Y|/(|Y| + |G|)$

Proof. Once again, relabel women so that w_i denotes the wife of m_i in $\mu_{\mathcal{M}}$. Say that a woman w_k with $k \neq i$ and to whom m_i prefers w_i is "red" if $k \leq i - C \ln N$, "yellow" if $i - C \ln N < k \leq i + 2C \ln N$, and "green" if $i + 2C \ln N < k$. Let R, Y and G be the sets of red, yellow and green women. Women who are not colored are ranked by m_i better than w_i, his best stable partner, so they cannot be stable partners of m_i.

If woman w_i has another stable partner besides m_i, then man m_i also has at least one other stable partner. Because of $\mathcal{K}$, every red woman w_k prefers m_k to m_i. Since m_k is her worst stable partner, there is no stable marriage in which m_i is paired with w_k. Thus all stable partners of m_i must be among $Y \cup G$.

Let w^* be m_i's favorite woman among $Y \cup G$. We will argue that if $w^* \in G$ then w_i is m_i's unique stable partner. Assume, for a contradiction, that m_i has another stable partner w besides w_i, and consider that stable matching μ. By Lemma 4.3, w must belong to i's tier. By $\mathcal{K}$ and since $w^* \in G$, w^* is in a different tier, so $w \neq w^*$. Consider the pair (m_i, w^*). By definition of w^*, man m_i prefers w^* to w. By $\mathcal{K}$ and definition of G, w^* prefers m_i to the man of her tier to whom she is married in μ. So (m_i, w^*) is a blocking pair, contradicting stability of μ. This proves

$$\mathbb{P}[w_i \text{ has more than 1 stable partner}] \leq \mathbb{P}[w^* \in Y].$$

Recall that m_i has popularity preferences. Once we condition on $\mathcal{H}$, the preferences of m_i are still popularity preferences over all the women to whom m_i prefers w_i. Event $\mathcal{K}$ only depends on the women's preference lists, so conditioning on $\mathcal{K}$ does not change that. Because popularities are bounded between 1 and $P \geq 1$, women in Y are at most P times more popular that women in G, and we have

$$\mathbb{P}[w^* \in Y] \leq \frac{|Y| \cdot P}{|Y| \cdot P + |G|} \leq \frac{3PC \ln N}{3PC \ln N + |G|}$$

where the second inequality comes from the fact that $|Y| \leq 3C \ln N$. Finally, we argue that all women w_j with $j > i + 2C \ln N$ are in G. Consider a woman w_j with $j > i + 2C \ln N$. Conditioning on $\mathcal{K}$, w_j prefers m_i to her partner w_j in $\mu_{\mathcal{M}}$, so by stability of $\mu_{\mathcal{M}}$, man m_i prefers w_i to w_j, so $w_j \in G$. Hence, conditioning on $\mathcal{H}$ and $\mathcal{K}$, we have $|G| = N - i - 2C \ln N$, and thus $3PC \ln N + |G| \geq N + C \ln N - i$. □

4.3.3 Putting everything together

Proof of Theorem 4.15. As in the previous proof, in our analysis we condition on event $\mathcal{H}$ (see Definition 4.7), i.e. on (1) the man-optimal stable matching $\mu_{\mathcal{M}}$, (2) each man's ranking of the women he prefers to his partner in $\mu_{\mathcal{M}}$, and (3) each woman's ranking of the men who prefer her to their partner in $\mu_{\mathcal{M}}$. As before, a person who is not matched in $\mu_{\mathcal{M}}$ remains single in all stable matchings, hence, without loss of generality, we assume that $M = W = N$, and that $w_i = \mu_{\mathcal{M}}(m_i)$ for all $1 \leq i \leq N$.

Let Z denote the number of women with several stable partners. We show that in expectation $Z = \mathcal{O}(\ln^2 N)$, hence the fraction of persons with multiple stable partners converges to 0. We separate the analysis of Z according to whether event $\mathcal{K}$ holds. When $\mathcal{K}$ does not hold, we bound that number by N, so by Lemma 4.17: $\mathbb{E}[Z] \leq (1/N^2) \times N + (1 - 1/N^2) \times \mathbb{E}(Z|\mathcal{K})$.
Conditioning on $\mathcal{H}$ and switching summations, we write:

$$\mathbb{E}(Z|\mathcal{K}) = \sum_{\mathcal{H}} \mathbb{P}[\mathcal{H}] \cdot \mathbb{E}(Z|\mathcal{K}, \mathcal{H}) = \sum_i \sum_{\mathcal{H}} \mathbb{P}[\mathcal{H}] \cdot \mathbb{P}[w_i \text{ has several stable husbands} \mid \mathcal{K}, \mathcal{H}]$$

By Lemma 4.18, we can write:

$$\mathbb{P}[w_i \text{ has several stable husbands} \mid \mathcal{K}, \mathcal{H}] \leq \frac{3PC \ln N}{N + C \ln N - i}.$$

Hence the expected number of women who have several stable partners is at most $1/N$ plus

$$\begin{aligned}
\sum_{i=1}^{N} \frac{3PC \ln N}{N + C \ln N - i} &= \sum_{i=0}^{N-1} \frac{3PC \ln N}{i + C \ln N} \\
&\leq 3PC \ln N \int_{C \log N - 1}^{C \log N - 1 + N} \frac{\mathrm{d}t}{t} \\
&= 3PC \ln N \ln \left(\frac{C \log N - 1 + N}{C \log N - 1} \right)
\end{aligned}$$

When N is large enough, we can simplify this bound to $3PC \ln^2 N$. □

4.4 Vertical - Vertical preferences

To complete the scenery, we study the case where both sides have strongly correlated preferences. This section serves as a transition towards the second part of this thesis: Who gets what?

We assume that men have aligned (Definition 3.10) popularity preferences (Definition 3.2), such that woman w_i has popularity $P(w_i) = \lambda^i$ with $0 < \lambda < 1$ and $1 \leq i \leq W$. Notice that it is a special case of conditionally monotone distribution, where the odds of woman w_{i+k} being ranked before woman w_i is equal to λ^k.

Theorem 4.19. *Assume that each woman independently draws her preference list from a conditionally-monotone distribution. Let u_k be an upper bound on the odds that man m_{i+k} is ranked before man m_i:*

$$\forall k \geq 1, \quad u_k = \max_{w,i} \left\{ \frac{\mathbb{P}[m_{i+k} \succ_w m_i]}{\mathbb{P}[m_i \succ_w m_{i+k}]} \;\middle|\; w \text{ finds both } m_i \text{ and } m_{i+k} \text{ acceptable} \right\}$$

Further assume that all preferences are complete, and that men have aligned popularity preferences where woman w_i has popularity $P(w_i) = \lambda^i$ with $0 < \lambda < 1$. Then for each man m_i, all of his stable partners are woman w_j where $|i - j| \leq \delta$, such that in expectation we have $\mathbb{E}[\delta] \leq 2\lambda/(1-\lambda)^3 + 2\exp(\sum_{k\geq 1} k u_k) \sum_{k \geq 1} k^2 u_k$.

The proof adapts the analysis from Section 4.2, and build tiers at the same time as the men-optimal stable matching $\mu_{\mathcal{M}}$. In Section 4.4.1, we run Algorithm 2.1 and decide to always pick the first single man in the order $m_1, \ldots, m_M$ to propose. Each time a woman receives her first proposal we check if all women who received offers prefer their current partners to men who have not proposed yet. If the answer is yes, we just found a prefix separator, and we can safely forget about men who already proposed and women who already received a proposal.

In Section 4.4.2 we use the fact that men have aligned popularity preferences: a crucial observation is that the order in which women receive their first proposal follows the popularity preferences distribution. Finally, we prove Theorem 4.19 in Section 4.4.3

4.4.1 Stochastic domination

Algorithm 4.2 Compute the interval of women with whom man m_n can be matched.

Initialize the set of super separators $S \leftarrow \{0\}$.
For i from 1 to M**, do**
  Set the proposer to be $m \leftarrow m_i$
  While the proposer is has not proposed to every woman**, do**
    m draws the next woman w_j in his preference list.
    If woman w_j is single**, then**
      Set $\sigma(i) \leftarrow j$, tentatively match m with w_j, and draw the preference list of w_j.
      If all women $w_{\sigma(1)}, \ldots w_{\sigma(i)}$ prefer their current partner to $m_{i+1}, \ldots, m_M$**, then**
        Add i to the set of super separators $S \leftarrow S \cup \{i\}$.
      Break out of the while loop.
    else if w_j prefer m to her current partner m' **then**
      Tentatively match w_j with m, and set the proposer to $m \leftarrow m'$.
Let l and r be two consecutive super separators such that $n \in (l, r]$.
Output the interval $[a, b]$, where $a \leftarrow \min([1, W] \setminus \sigma([1, l]))$ and $b \leftarrow \max(\sigma([1, r]))$.

Lemma 4.20. *Given n, Algorithm 4.2 computes an interval $[a, b]$ such that all the stable wifes of man m_n belong to the set $\{w_a, w_{a+1}, \dots, w_b\}$.*

Proof. We show that super-separators (in Algorithm 4.2) are separators (see Definition 4.4): if at some iteration $1 \leq i \leq M$ all women prefer their current partner to $m_{i+1}, \dots, m_M$, then their current partner is the one from $\mu_{\mathcal{M}}$. Remark that not all separators are super-separators. Two consecutive super-separators l and r such that $n \in (l, r]$ define a super-tier, which is the union of (possibly) several consecutive tiers. Thus, all stable partners of man m_n belong to the set $\{w_{\sigma(l+1)}, \dots, w_{\sigma(r)}\}$. Let a be minimal such that woman w_a does not belong to $\{w_{\sigma(1)}, \dots, w_{\sigma(l)}\}$, and let b be maximal such that w_b belongs to $\{w_{\sigma(1)}, \dots, w_{\sigma(r)}\}$. By construction, all the stable wifes of man m_n belong to the set $\{w_a, \dots, w_b\}$. □

Lemma 4.21. *Given n, the super-separators l and r defined by Algorithm 4.2 are such that $r - n$ and $n - 1 - l$ are both stochastically dominated by X from Definition 4.11.*

Proof. To prove that X dominates $r - n$, we adapt the proof of Lemma 4.12. The way we compute the super-separator r is nearly identical to the left-to-right greedy method from Algorithm 4.1. Starting from $r_0 = n$, we let r_1 be the maximum i for which a woman $w \in \{w_{\sigma(1)}, \dots, w_{\sigma(n)}\}$ prefer m_i to her current partner. The jump $r_1 - r_0$ is dominated by the random variable Δ_0 from Definition 4.11. We define the sequence $(r_t)_{t \geq 1}$ by induction, and bound each jump $r_{t+1} - r_t$ by Δ_t. This concludes the proof that X dominates $r - n$.

To prove that X dominates $n - 1 - l$, we make the following observation: if $n - 1 - l > \delta$ for some constant δ, then $r \geq n - \delta - 1 > l$ and $n' = n - \delta - 1$ belongs to the same super-tier as n. In particular we have $r - n' > \delta$, which happens with probability smaller that $\mathbb{P}[X > \delta]$ using the stochastic domination of $r - n'$ by X. □

4.4.2 Sequence of proposals

Recall that we assume that men have identical popularity preferences over women, such that woman w_i has popularity $P(w_i) = \lambda^i$ with $0 < \lambda < 1$ and $1 \leq i \leq W$. When running Algorithm 4.2, we sample preferences of men online, and record each time a proposal is made to a woman for the first time. Lemma 4.22 shows that the random ordering induced by first proposals has exactly the same distribution as the popularity preferences induced by P. Lemma 4.23 gives an upper bound on the probability of rare events.

Lemma 4.22. *Assuming that all men have aligned (Definition 3.10) complete popularity preferences (Definition 3.2) induced by $P : \mathcal{W} \to \mathbb{R}_+^*$. The ordering $w_{\sigma(1)} \succ \dots \succ w_{\sigma(M)}$ defined in Algorithm 4.2 has the same distribution as the popularity preferences.*

Proof. The ordering σ is defined by the order in which women receive their first proposal. In particular, if we condition on the beginning $\sigma(1), \dots, \sigma(k)$ of the ordering, then $\sigma(k+1)$ is drawn at random from $[1, W] \setminus \sigma([1, k])$, and is equal to j with probability proportional to $P(w_j)$. This stochastic process is identical to the one used to build popularity preferences. □

Lemma 4.23. *Consider complete popularity preferences, such that each woman w_i has popu-*

larity $P(w_i) = \lambda^i$ with $0 < \lambda < 1$ and $1 \le i \le W$. For every $1 \le k \le W$ and $\delta > 1$,

$$\mathbb{P}[\exists i \le k - \delta,\ \textit{woman } w_i \textit{ is not ranked in the top } k] \le \frac{\lambda^{\delta+1}}{(1-\lambda)^2} = \exp(-\Omega(\delta))$$

$$\mathbb{P}[\exists i > k + \delta,\ \textit{woman } w_i \textit{ is ranked in the top } k] \le \frac{\lambda^{\delta+1}}{(1-\lambda)^2} = \exp(-\Omega(\delta))$$

Proof. Let $i \le k - \delta$. If w_i is not ranked in the top k, it means that at least one woman among $w_{k+1}, \dots, w_W$ is ranked in the top k, and thus that w_i is not ranked before all of them.

$$\begin{aligned}\mathbb{P}[w_i \text{ is not ranked in the top } k] &\le 1 - \mathbb{P}[w_i \succ w_{k+1}, \dots, w_W] \\ &= 1 - \frac{\lambda^i}{\lambda^i + \sum_{j=k+1}^{W} \lambda^j} \\ &\le \frac{\sum_{j=k+1}^{W} \lambda^j}{\lambda^i} \le \frac{\lambda^{k+1-i}}{1-\lambda}\end{aligned}$$

Using the union-bound, we sum the upper bound for every $i \le k - \delta$. We obtain

$$\mathbb{P}[\exists i \le k - \delta,\ \text{woman } w_i \text{ is not ranked in the top } k] \le \sum_{i=1}^{k-\delta} \frac{\lambda^{k+1-i}}{1-\lambda} \le \frac{\lambda^{\delta+1}}{(1-\lambda)^2}$$

Let $i > k + \delta$. If w_i is ranked in the top k, it means that at least one woman among $w_1, \dots, w_k$ is not ranked in the top k, and thus that w_i is ranked before one of $w_1, \dots, w_k$.

$$\begin{aligned}\mathbb{P}[w_i \text{ is ranked in the top } k] &\le \mathbb{P}[w_i \text{ is ranked before one of } w_1, \dots, w_k] \\ &\le \sum_{j=1}^{k} \mathbb{P}[w_i \succ w_j] = \sum_{j=1}^{k} \frac{\lambda^i}{\lambda^i + \lambda^j} \le \frac{\lambda^{i-k}}{1-\lambda}\end{aligned}$$

Using the union-bound, we sum the upper bound for every $i > k + \delta$. We obtain

$$\mathbb{P}[\exists i > k + \delta,\ \text{woman } w_i \text{ is ranked in the top } k] \le \sum_{i=k+\delta+1}^{W} \frac{\lambda^{i-k}}{1-\lambda} \le \frac{\lambda^{\delta+1}}{(1-\lambda)^2}$$

□

4.4.3 Putting everything together

Proof of Theorem 4.19. Given n, we use Algorithm 4.2 to compute the interval $[a, b]$ in which all stable partners of m_n belong.

First, we use Lemma 4.21 to show that both $r-n$ and $n-1-l$ are stochastically dominated by X. Hence, in Algorithm 4.2 we can replace the definition of r and l by $r' = n - x_1$ and $l' = n - 1 - x_2$ where x_1 and x_2 are drawn from the same distribution as X. The modified algorithm gives an interval $[a', b']$ which stochastically dominates the original interval (that is, a dominates a', and b' dominates b).

Second, we change the order in which Algorithm 4.2 computes σ, l' and r'. Because the random variables l' and r' are independent from σ, they can be computed at the beginning. Conditioning on the values of l' and r', we use Lemmas 4.22 and 4.23 to bound a' and b'.

$$\mathbb{E}[b' - r'] = \sum_{\delta \geq 0} P[b' - r' > \delta] \leq \sum_{\delta \geq 0} \frac{\lambda^{\delta+1}}{(1-\lambda)^2} \leq \frac{\lambda}{(1-\lambda)^3}$$
$$\mathbb{E}[l' - a' + 1] = \sum_{\delta \geq 0} P[l' - a' \geq \delta] \leq \sum_{\delta \geq 0} \frac{\lambda^{\delta+1}}{(1-\lambda)^2} \leq \frac{\lambda}{(1-\lambda)^3}$$

Finally, we write $b' - n = (b' - r') + x_1$ and $n - a' = (l' - a' + 1) + x_2$. □

4.5 Simulations

Implementations of two-sided matching markets with popularity preferences are shared between Chapters 4, 8 and 5, and are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/Popularity

For simulation purposes, consider that women have aligned popularity preferences induced by $P : m_i \mapsto 2^{-i}$. The odds of man m_{i+k} being ranked before man m_i is exactly equal to $u_k = 2^{-k}$. In Figures 4.5, 4.6 and 4.7, men have random preferences ranging from horizontal to vertical. Figure 4.5 plots the match probabilities, Figure 4.6 plots the expected ranks of each person's worst and best stable partner, and Figure 4.7 plots the number of stable partners of each person.

Using Theorem 4.1, the expected rank difference between a woman's worst and best stable partner is at most $(1 + 2\exp(\sum_{k\geq 1} k \cdot u_k) \sum_{k\geq 1} k^2 \cdot u_k$, which is equal to $(1 + 2e^2) \cdot 6 \approx 95$ when $u_k = 2^{-k}$. Figures 4.6 and 4.7 confirms this result, and even shows that our upper-bound can be improved when preferences of men are not adversarial. Using Theorem 4.15, the fraction of women who have multiple stable partners tends towards 0 when men have uniform preferences. In Figure 4.7, panel (c) confirms this result, and panel (a) shows that the hypothesis on the preferences of men is necessary. Using Theorem 4.19, stable matchings are assortative when men and women have vertical preferences. More precisely, we can bound the difference of index $|i - j|$ between a man m_i and a woman w_j who are matched in a stable matching. Figure 4.6 confirms this result, and show that our upper-bound can be improved.

Finally, simulations illustrate the intuitive explanation of our results in terms of imbalance that we sketched in the introduction. When the preferences of agents from one side of the market are strongly correlated, it creates small "tiers" of agents on the opposite side. When the correlations between preferences of women are stronger than correlation between preferences of men, tiers of men are smaller than tiers of women, which creates an imbalance. Any imbalance is favorable for the agents in the small side of the market, who have a larger set of choices for partners. This explains why men give smaller rank to their stable partners in Figure 4.6.

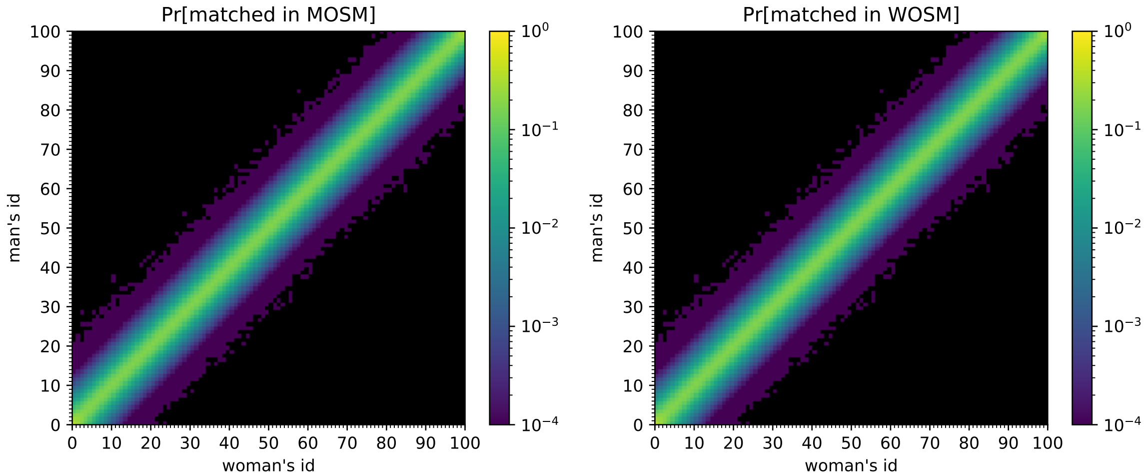
4.6 Conclusion and open questions

In this chapter, we explore the effect of correlated preferences on the core-convergence phenomenon. We show that the effect is strongest in the vertical-horizontal case, that is when correlations induce an imbalance in the market. The following questions are left open for future work:

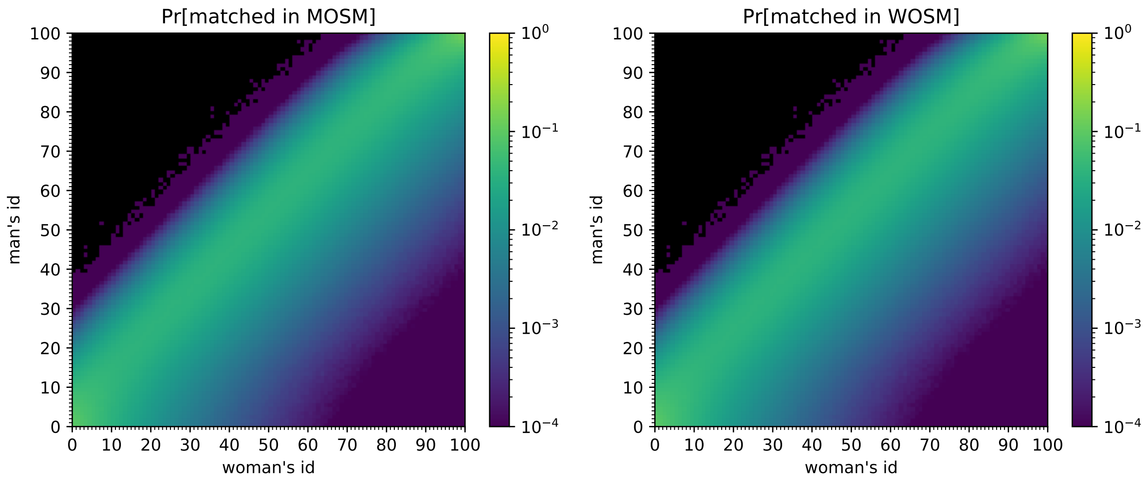
- **Deterministic core-convergence.** As discussed in the introduction, Holzman and Samet [HS14] give deterministic analogues of Theorems 4.1 and 4.19. Finding sufficient conditions for a deterministic version of Theorem 4.15 would be an interesting result.

- **Rank correlation coefficients.** In statistics, correlations between rankings are usually measured by distances such as Kendall's τ and Spearman's ρ coefficients. In Mallow's model, rankings are drawn with probability exponentially decreasing in Kendall's τ distance with a base ranking. Unfortunately, such distributions are not conditionally-monotone and thus our proof techniques do not apply. However, it would be interesting to understand the relation between the value of a correlation coefficient and the core-convergence phenomenon. Computer simulations from Celik and Knoblauch [CK07] make a first step in this direction.

(a) Men have “vertical” preferences: popularity distribution with $P : w_i \mapsto 0.5^i$.



(b) Men have “vertical” preferences: popularity distribution with $P : w_i \mapsto 0.9^i$.



(c) Men have “horizontal” preferences: uniform distribution.

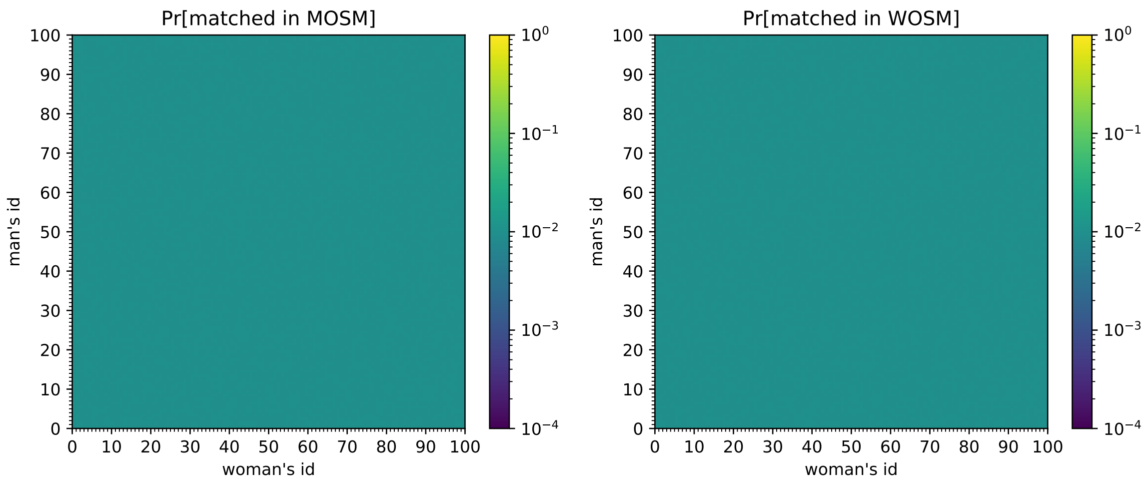
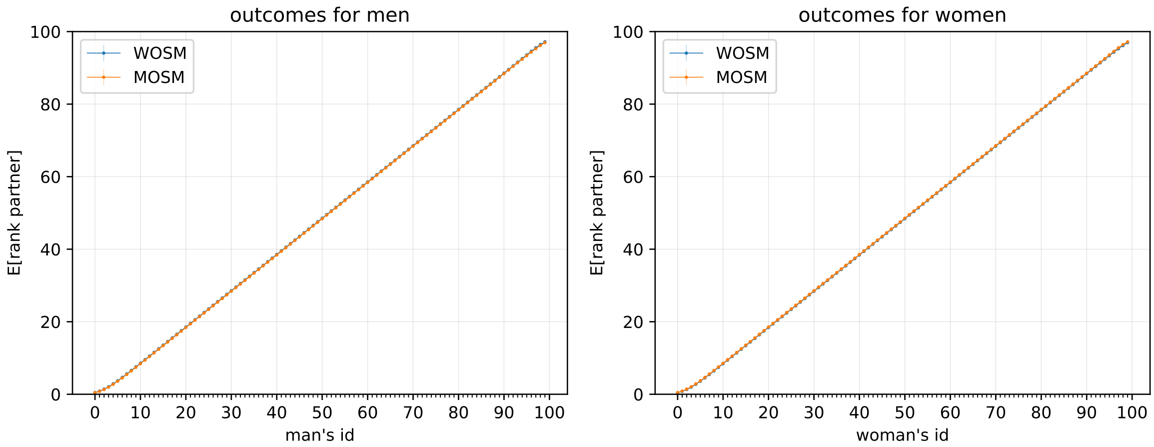
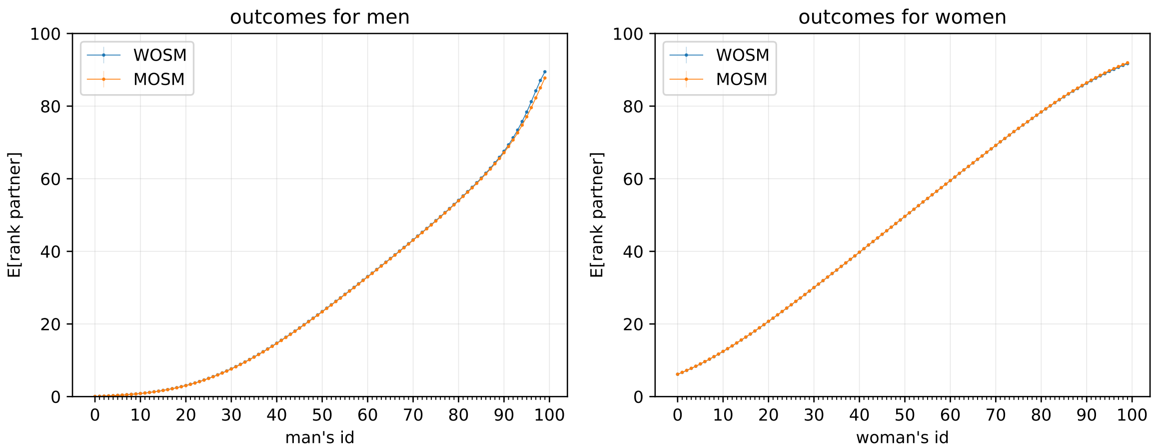


Figure 4.5. Probability of a man and a woman being matched, under the men-optimal-stable-matching (MOSM) and the women-optimal-stable-matching (WOSM). Women have “vertical” preferences, built using the popularities distribution $P : m_i \mapsto 0.5^i$. Plots contain the average values over 10^6 runs, with 100 men and 100 women.

(a) Men have "vertical" preferences: aligned popularity $P : w_i \mapsto 0.5^i$ distribution.



(b) Men have "vertical" preferences: aligned popularity $P : w_i \mapsto 0.9^i$ distribution.



(c) Men have "horizontal" preferences: uniform distribution.

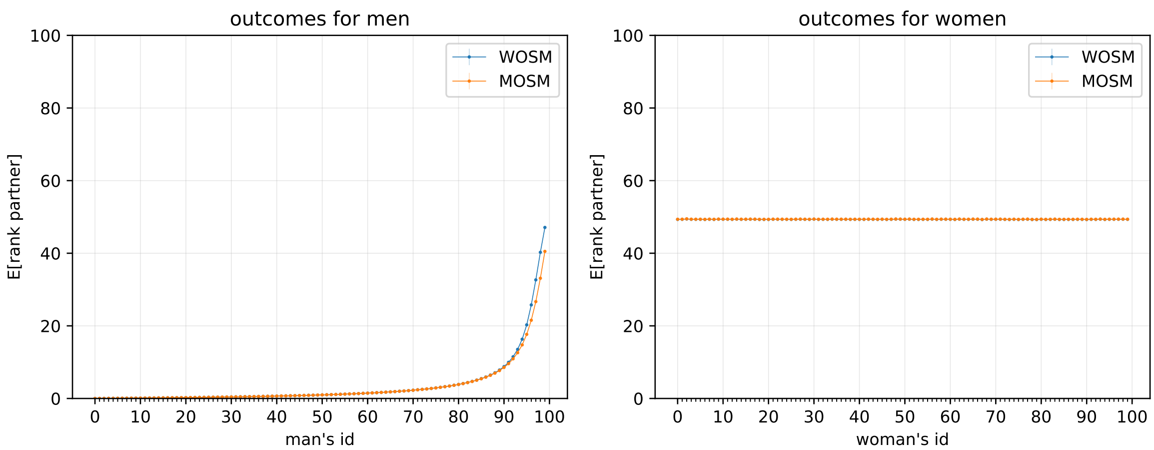
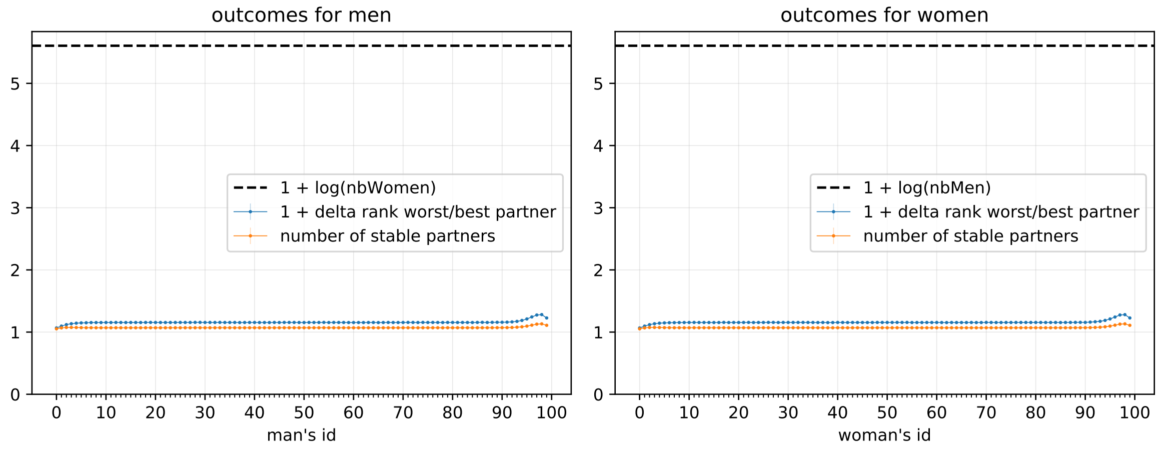
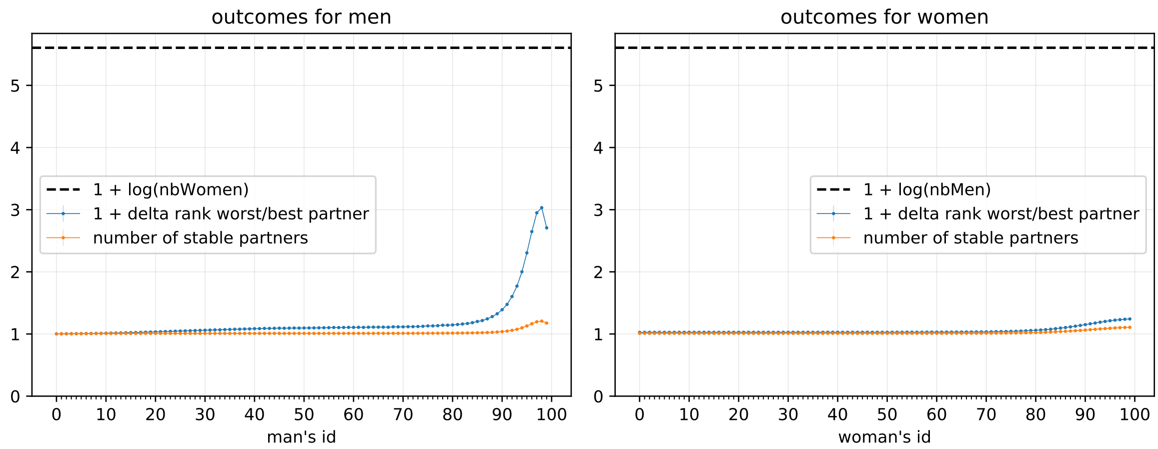


Figure 4.6. Expected rank of each person's partner, under the men-optimal-stable-matching (MOSM) and the women-optimal-stable-matching (WOSM). Women have "vertical" preferences, built using the popularities distribution $P : m_i \mapsto 0.5^i$. Plots contain the average values over 10^6 runs, with 100 men and 100 women.

(a) Men have “vertical” preferences: popularity distribution with $P : w_i \mapsto 0.5^i$.



(b) Men have “vertical” preferences: popularity distribution with $P : w_i \mapsto 0.9^i$.



(c) Men have “horizontal” preferences: uniform distribution.

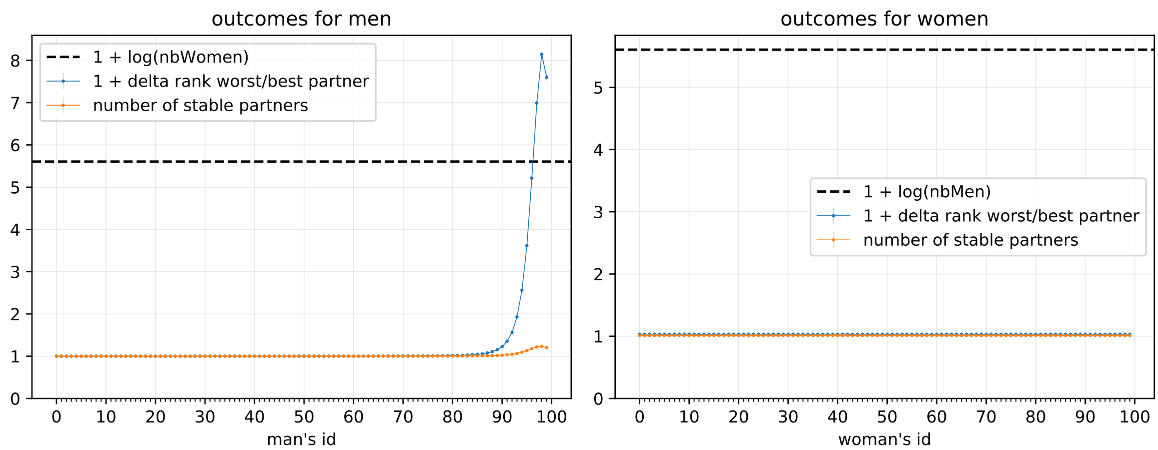


Figure 4.7. Number of stable partners of each person. Women have “vertical” preferences, built using the popularities distribution $P : m_i \mapsto 0.5^i$. Plots contain the average values over 10^6 runs, with 100 men and 100 women.

5 | Counting Stable Pairs

This chapter is based on the following paper:

[GMM19] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Two-Sided Matching Markets with Correlated Random Preferences Have Few Stable Pairs". In: *arXiv preprint arXiv:1904.03890* (2019)

5.1 Introduction

As discussed in Chapter 2, matching markets with a unique stable matching are incentive compatible, in the sense it is a dominant strategy for every agent to be truthful. Empirically, researchers have observed that in practice, there is often a nearly unique stable matching. This is the *core-convergence* phenomenon, which gives weaker notions of strategy-proofness where the incentives of agents to manipulate are limited. In Chapter 4, we bound the fraction of agents who have multiple stable partners and the difference in rank/utility between someone's worst and best stable partner.

As an attempt to model real matching markets, a series of papers [Pit89; KMP90; Pit92; PSV07; LP09] study the setting where N men and N women have complete random uniform preferences over one another: the expected rank of each person's best and worst partners are respectively $\sim \ln N$ and $\sim N/\ln N$. In such situations, measuring the number of agents with multiple stable partners and the rank difference between those partners is insufficient, and it is interesting to look at alternative measures: the expected number of stable matchings is $\sim e^{-1} N \ln N$, and the expected total number of stable pairs is $\sim N \ln N$.

In order to quantify "weak core-convergence", we argue that counting the total number of stable pairs is a good measure, whereas counting the number of stable matchings is not. Figure 5.1 gives three instances of stable matchings with N men and N women. We observe that the number of stable matchings can be arbitrarily large, because of multiple independent sub-markets having multiple stable matchings. Conversely, the total number of stable pairs has the nice property of measuring how different stable matchings are. This motivates our decision to study the number of stable pairs, even though this measure is not directly related to incentive compatibility.

Theorem 5.1. *Assume that either (1) or (2) holds:*

(1) Women have aligned popularity preferences,

(2) Men and women have symmetric popularity preferences.

Then expected number of stable pairs is at most $N(1 + \ln N)$.

In this chapter, we give upper bounds on the total number of stable pairs when preference are drawn from popularity distributions (see Definition 3.2). In particular, when agents have aligned (see Definition 3.10) or symmetric (see Definition 3.11) popularity preferences, Theorem 5.1 shows

(a) The preferences of agents are almost aligned, with several local swaps. Each person has 2 stable partners, and the market can be decomposed into $N/2$ independent sub-markets, each having 2 stable matchings.

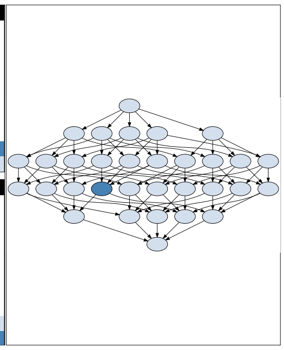
Nb. stable pairs $= 2N$

Nb. stable matchings $= 2^{N/2}$

1	2	3	4	5	6	7	8	9	10
A	B	A	A	A	A	A	A	A	A
B	A	B	B	B	B	B	B	B	B
C	C	C	D	C	C	C	C	C	C
D	D	D	C	D	D	D	D	D	D
E	E	E	E	E	F	E	E	E	E
F	F	F	F	F	E	F	F	F	F
G	G	G	G	G	G	G	H	G	G
H	H	H	H	H	H	H	G	H	H
I	I	I	I	I	I	I	I	I	J
J	J	J	J	J	J	J	J	J	I

A	B	C	D	E	F	G	H	I	J
2	1	1	1	1	1	1	1	1	1
1	2	2	2	2	2	2	2	2	2
3	3	4	3	3	3	3	3	3	3
4	4	3	4	4	4	4	4	4	4
5	5	5	5	6	5	5	5	5	5
6	6	6	6	5	6	6	6	6	6
7	7	7	7	7	7	8	7	7	7
8	8	8	8	8	8	7	8	8	8
9	9	9	9	9	9	9	9	10	9
10	10	10	10	10	10	10	10	9	10

Preference lists



Lattice of stable matchings

(b) The preferences of agents have been drawn uniformly at random.

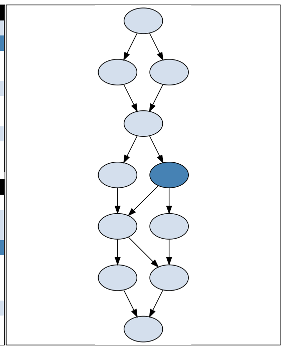
$\mathbb{E}$[Nb. stable pairs] $\sim N \ln N$

$\mathbb{E}$[Nb. stable matchings] $\sim e^{-1} N \ln N$

1	2	3	4	5	6	7	8	9	10
D	H	A	C	H	F	C	E	F	I
G	G	E	G	F	D	G	B	J	J
J	C	I	A	J	J	B	J	G	F
H	I	G	E	G	H	J	G	C	B
B	E	H	H	C	B	D	A	B	D
F	A	B	F	I	E	A	C	E	A
C	F	D	B	B	G	I	I	D	H
E	D	C	I	A	C	E	H	A	C
I	B	F	J	D	I	F	D	H	G
A	J	J	D	E	A	H	F	I	E

A	B	C	D	E	F	G	H	I	J
3	4	3	2	9	8	6	8	9	3
4	8	10	4	3	4	8	1	1	9
5	5	1	9	1	9	9	2	4	1
2	7	4	10	10	10	7	3	7	10
7	10	7	1	5	6	3	10	2	8
1	2	8	3	8	2	5	4	6	4
8	9	6	7	7	7	10	9	8	7
6	1	9	5	2	3	2	6	5	5
9	3	2	6	4	1	4	5	3	2
10	6	5	8	6	5	1	7	10	6

Preference lists



Lattice of stable matchings

(c) The preferences of agents are cyclic, such that for each man m and woman w the sum of the ranks of m in w's list and w in m's list is equal to $N+1$.

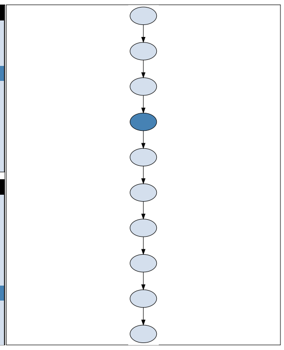
Nb. stable pairs $= N^2$

Nb. stable matchings $= N$

1	2	3	4	5	6	7	8	9	10
A	B	C	D	E	F	G	H	I	J
B	C	D	E	F	G	H	I	J	A
C	D	E	F	G	H	I	J	A	B
D	E	F	G	H	I	J	A	B	C
E	F	G	H	I	J	A	B	C	D
F	G	H	I	J	A	B	C	D	E
G	H	I	J	A	B	C	D	E	F
H	I	J	A	B	C	D	E	F	G
I	J	A	B	C	D	E	F	G	H
J	A	B	C	D	E	F	G	H	I

A	B	C	D	E	F	G	H	I	J
2	3	4	5	6	7	8	9	10	1
3	4	5	6	7	8	9	10	1	2
4	5	6	7	8	9	10	1	2	3
5	6	7	8	9	10	1	2	3	4
6	7	8	9	10	1	2	3	4	5
7	8	9	10	1	2	3	4	5	6
8	9	10	1	2	3	4	5	6	7
9	10	1	2	3	4	5	6	7	8
10	1	2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8	9	10

Preference lists



Lattice of stable matchings

Figure 5.1. Three instances with $N = 10$ men and $N = 10$ women having complete preferences. From top to bottom, the number of stable matchings (on the right) is decreasing and the number of stable pairs (in blue, on the left) is increasing. We argue that weak core-convergence is best captured by a small number of stable pairs.

that the expected total number of stable pairs is at most $N(1+\ln N)$, which matches the asymptotic value when agents have uniform preferences. Our main technical result is Theorem 5.4, which bound the expected number of stable husbands of a woman having popularity preferences, via a stochastic analysis of Algorithm 2.2, which generalize the upper bound from [KMP90]. We prove Theorem 5.1 in Section 5.5 and we give high probability bounds in Section 5.6. To illustrate our results, we provide a tight example in Section 5.4 and we present experimental results in Section 5.7.

Related works. Wilson [Wil72] and Knuth [Knu76] show that when men have uniform preferences, the average complexity of the men proposing deferred acceptance procedure is $\sim N \log N$, which implies that the expected rank of each person's best stable partner is $\sim \ln N$. Pittel [Pit89] show that when both men and women have uniform preferences, the expected rank of each person's worst stable partner is $N/\ln N$, and that the expected number of stable matchings is $\sim e^{-1} N \ln N$. Knuth, Motwanni and Pittel [KMP90; Pit92] show that the expected number of stable husbands of each woman is $\sim \ln N$, where the upper bound holds when women have uniform preferences and the lower bound holds when both men and women have uniform preferences.

Takeaway message. We argue that counting stable pairs quantifies "weak core-convergence" when most agents do not have a unique stable partner but the situation is still far from being worst case. To give an intuitive interpretation of our results, one can examine Figure 5.1: in panel (a), preferences of agents are positively correlated (almost aligned), and the number of stable pairs is low; in panel (b), preferences are uncorrelated (uniformly distributed), and the number of stable pairs is intermediate; in panel (c), preferences are negatively correlated (cyclic and reversed), and the number of stable pairs is high. We model positive correlations using symmetric or aligned popularity preferences. Theorem 5.1 show that the uniform case is a worst case situation when compared to the positively correlated case. In a related paper, Boudreau and Knoblauch [BK10] discuss positive and negative "intercorrelations" of preferences and their effect on the overall welfare.

5.2 Popularity preferences

Recall that in our model for probability preferences, the set of acceptable partners is deterministic. To build her preference list, w samples without replacement from her set of acceptable men, first drawing her favourite partner, then her second favourite, and so on until her least favourite acceptable partner.

Lemma 5.2. *Assume that a woman w has popularity preferences defined by P_w. Conditioning on a partial ranking of acceptable men $a_1 \succ_w \cdots \succ_w a_p$, the probability that w rank m before a_1 is exactly*

$$\mathbb{P}[m \succ_w a_1 \mid a_1 \succ_w \cdots \succ_w a_p] = \frac{P_w(m)}{P_w(m) + \sum_{i=1}^{p} P_w(a_i)}$$

Proof. One nice feature of popularity preferences is that to compute the probability that $a_1 \succ_w \cdots \succ_w a_p$, one can ignore each time a man not in $\{a_1, \ldots, a_p\}$ is drawn. We obtain

$$\mathbb{P}[a_1 \succ_w \cdots \succ_w a_p] = \prod_{i=1}^{p} \frac{P_w(a_i)}{\sum_{j=i}^{p} P_w(a_j)}.$$

and similarly for the probability that $m \succ_w a_1 \succ_w \cdots \succ_w a_p$. Thus,

$$\mathbb{P}[m \succ_w a_1 \mid a_1 \succ_w \cdots \succ_w a_p \succ_w w] = \frac{\mathbb{P}[m \succ_w a_1 \succ_w \cdots \succ_w a_p \succ_w w]}{\mathbb{P}[a_1 \succ_w \cdots \succ_w a_p \succ_w w]} = \frac{P_w(m)}{P_w(m) + \sum_{i=1}^{p} P_w(a_i)}$$

□

As in Chapter 4, we will use the *principle of deferred decision* and construct preference lists in an online manner. By Theorem 2.3 the man-optimal stable matching $\mu_{\mathcal{M}}$ is computed by Algorithm 2.1, and the remaining randomness can be used for a stochastic analysis of each person's stable partners. To be more formal, we define a random variable $\mathcal{H}$, and inspection of Algorithm 2.1 shows that $\mathcal{H}$ contains enough information on each person's preferences to run Algorithm 2.1 deterministically.

Definition 5.3. Let $\mathcal{H} = (\mu_{\mathcal{M}}, (\sigma_m)_{m\in\mathcal{M}}, (\pi_w)_{w\in\mathcal{W}})$ denote the random variable consisting of (1) the man-optimal stable matching $\mu_{\mathcal{M}}$, (2) each man's ranking of the women he prefers to his partner in $\mu_{\mathcal{M}}$, and (3) each woman's ranking of the men who prefer her to their partner in $\mu_{\mathcal{M}}$.

5.3 One person has popularity preferences

Theorem 5.4. *Let w be a woman. Assume that w has popularity preferences defined by P_w and that she has at least one stable partner. The preference lists of the men and of the women other than w are arbitrary. Then*

$$\mathbb{E}[\textit{Number of stable husbands of } w] \leq 1 + \ln d_w + \mathbb{E}\left[\ln \frac{P_w(\mu_{\mathcal{W}}(w))}{P_w(\mu_{\mathcal{M}}(w))}\right],$$

where d_w denotes the number of acceptable husbands of w, $\mu_{\mathcal{M}}(w)$ is her worst stable partner and $\mu_{\mathcal{W}}(w)$ is her best stable partner.

Proof. First, observe that w is matched if and only if she receives a proposal in Algorithm 2.1, which is independent from her ordering of acceptable men. By Theorem 2.5, Algorithm 2.2 outputs the stable husbands of w, so we analyze that algorithm, which starts by a call to Algorithm 2.1, which by Theorem 2.3 yields matching $\mu_{\mathcal{M}}$.

We know the preferences of everyone except w. We start the analysis by conditioning on the random variable $\mathcal{H}$, i.e. on woman w's ranking of the men who prefer her to $m_{\mathcal{M}}(w)$ (see Definition 5.3). From here, observe that the execution of Algorithm 2.1, and of the first phase of Algorithm 2.2 are deterministic. Let $x_0 = \mu_{\mathcal{M}}(w)$ be w's worst stable husband, K denote the number of proposals received by w in Phase 1, and let $x_1, x_2, \ldots, x_K$ denote the sequence of proposals received by w during the first phase of Algorithm 2.2.

Let p denote the sum of popularities of proposals received by w during the initial call to Algorithm 2.1, including $\mu_{\mathcal{M}}(w)$, and let $p_i = P_w(x_i)$ for all $0 \leq i \leq K$. By linearity of expectations, and then using Lemma 5.2,

$$\begin{aligned} \mathbb{E}[\text{Nb of stable husbands of } w \mid \mathcal{H}] &= 1 + \textstyle\sum_{i=1}^{K} \mathbb{P}[\text{proposal } x_i \text{ is accepted by } w \mid \mathcal{H}] \quad (5.1)\\ &= 1 + \textstyle\sum_{i=1}^{K} p_i/(p + p_1 + \cdots + p_i). \quad (5.2) \end{aligned}$$

We simplify the right-hand side with a sum-integral comparison:

$$\sum_{i=1}^{K} \frac{p_i}{p + p_1 + \cdots + p_i} \leq \sum_{i=1}^{K} \int_{p+p_1+\cdots+p_{i-1}}^{p+p_1+\cdots+p_i} \frac{\mathrm{d}t}{t} = \ln(p + p_1 + \cdots + p_K) - \ln p. \quad (5.3)$$

We use the convexity of $t \mapsto t \ln t$ and Jensen's inequality:

$$\ln(p + p_1 + \cdots + p_K) \leq \ln(K+1) + \frac{p \ln p + p_1 \ln p_1 + \cdots + p_K \ln p_K}{p + p_1 + \cdots + p_K}. \quad (5.4)$$

We now focus on the right-hand side of the equation in Lemma 5.2. By definition of $\mu_{\mathcal{W}}$, man $\mu_{\mathcal{W}}(w)$ is the overall best proposition received by w. It is x_0 with probability proportional to p and it is x_i $(1 \le i \le K)$ with probability proportional to p_i, thus

$$\mathbb{E}\left[\ln(P_w(\mu_{\mathcal{W}}(w))) \,|\, \mathcal{H}\right] = \frac{p \ln p_0 + p_1 \ln p_1 + \cdots + p_K \ln p_K}{p + p_1 + \cdots + p_K}. \tag{5.5}$$

Combining Equations (5.2), (5.3), (5.4) and (5.5), we obtain

$$\mathbb{E}[\text{Nb of stable husbands of } w] \le 1 + \ln(K+1) + \mathbb{E}[\ln(P(\mu_{\mathcal{W}}(w^*)))] - \frac{p \ln p_0 - p \ln p}{p + p_1 + \cdots + p_K} - \ln p,$$

where all expectations are conditioned on $\mathcal{H}$. Since $p \ge p_0$, we can write

$$\frac{p \ln p_0 - p \ln p}{p + p_1 + \cdots + p_K} + \ln p = \frac{p \ln p_0 + (p_1 + \ldots + p_K) \ln p}{p + p_1 + \cdots + p_K} \ge \ln p_0 = \ln P_w(\mu_{\mathcal{M}}(w)).$$

To conclude the proof, observe that $K + 1 \le d_w$ and take expectations over $\mathcal{H}$. □

5.4 Tight example for the number of stable partners

Knuth, Motwanni and Pittel [KMP90; Pit92] proved that when all persons have complete uniform preference lists each person has asymptotically $\ln N$ stable partner, and thus the upper-bound of Theorem 5.4 is tight. Here, we give another example showing that the upper bound from Theorem 5.4 is also tight when w has complete popularity preferences $P_w : m_i \mapsto \lambda^i$ with parameter $\lambda = 0.99$.

In the upper bound from Theorem 5.4, the ratio of popularity is at most λ^{1-M}, hence its logarithm is at most $(1-M)\ln\lambda$. When $\lambda = 0.99$, we get $(1-M)\ln\lambda \approx (1-\lambda)\cdot M$, hence Theorem 5.4 states that at most $\approx 1\%$ of the men are stable husbands of w. Lemma 5.5 proves that there exists an instance such that this 1% upper bound is asymptotically tight.

Lemma 5.5. *Let w a woman having complete popularity preferences $P_w : m_i \mapsto \lambda^i$ with $0 < \lambda < 1$. One can choose the preference lists of the other persons such that:*

$$\mathbb{E}[\textit{Number of stable husbands of } w] > (1-\lambda)\cdot M$$

Proof. Take a community with N men and N women. We adapt a folklore instance where each man-woman pair is stable (see Figure 5.2). We replace the preference list of woman w_1 by a complete popularity preference list defined by $P_w : m_i \mapsto \lambda^i$ with $0 < \lambda < 1$, which tends to be to similar with the original preference list $m_1 \succ m_2 \succ \cdots \succ m_N$.

During the execution of Algorithm 2.1, each man proposes to his favorite woman, and each woman receives exactly one proposition. Thus in the man-optimal stable matching, we have $\mu_{\mathcal{M}}(w_1) = m_N$, $\mu_{\mathcal{M}}(w_2) = m_1$, …, $\mu_{\mathcal{M}}(w_N) = m_{N-1}$.

Then, during the execution of Algorithm 2.2, woman $w^* = w_1$ will receive propositions from m_{N-1}, …, m_2, m_1, exactly in that order. The proposal from man m_i will be the "best so far" with probability $\lambda^i / \sum_{j=i}^{N} \lambda^j$. Hence, w^*'s expected number of stable husbands is exactly

$$\mathbb{E}[\text{Number of stable husbands of } w] = \sum_{i=1}^{N} \frac{\lambda^i}{\sum_{j=i}^{N} \lambda^j} = \sum_{i=1}^{N} \frac{1-\lambda}{1-\lambda^{N-i+1}} > (1-\lambda)\cdot N$$

□

m_1	$w_2 \succ w_3 \succ \dots \succ w_{N-1} \succ w_N \succ w_1$
m_2	$w_3 \succ \dots \succ w_{N-1} \succ w_N \succ w_1 \succ w_2$
$\vdots$	$\vdots$
m_{N-1}	$w_N \succ w_1 \succ w_2 \succ w_3 \succ \dots \succ w_{N-1}$
m_N	$w_1 \succ w_2 \succ w_3 \succ \dots \succ w_{N-1} \succ w_N$

w_1	complete popularity preferences $P_w : m_i \mapsto \lambda^i$, for some $0 < \lambda < 1$.
w_2	$m_2 \succ \dots \succ m_{N-2} \succ m_{N-1} \succ m_N \succ m_1$
$\vdots$	$\vdots$
w_{N-1}	$m_{N-1} \succ m_N \succ m_1 \succ m_2 \succ \dots \succ m_{N-2}$
w_N	$m_N \succ m_1 \succ m_2 \succ \dots \succ m_{N-2} \succ m_{N-1}$

Figure 5.2. Instance where woman w_1 has around $(1 - \lambda) \cdot M$ stable partners.

5.5 Correlated preferences

Intrinsic popularities model "one-sided" correlations, for example when all women agree that some men are more popular. Symmetric popularities model "cross-sided" correlations, for example when men and women prefer partners with whom they share some centers of interest. Both intrinsic popularities and symmetric popularities generalizes the uniform case. The upper bound from Theorem 5.1 matches the bound from [Pit92], implying that uncorrelated preferences are a worst case situation up to lower order terms: correlations reduce the number of stable pairs.

5.5.1 Aligned preferences

In this subsection, all women have popularity preferences. We say that men have *intrinsic* popularities when all women agree on the popularity of each man they find acceptable. To measure the extent to which women agree on the popularity of men, we introduce a parameter $r_m \geq 1$ which is the ratio between the highest and the lowest popularity given to man m by some woman who finds him acceptable. Lower values of r_m's mean more correlations between the preferences of women. When $r_m = 1$, all women agree on the intrinsic popularity of man m.

Theorem 5.6. *Assume that each woman w has popularity preferences defined by P_w over a set of $d_w \geq 1$ acceptable men. For each man m, we define the ratio r_m between the highest and the lowest popularity given to m by some woman who finds him acceptable. Then:*

$$\mathbb{E}[\textit{Number of stable pairs}] \leq N + \sum_{w \in \mathcal{W}} \ln d_w + \sum_{m \in \mathcal{M}} \ln r_m$$

When all women agree on the popularity of men, the number of stable pairs is at most $N + N \ln N$.

Proof. Theorem 5.4 is valid for each woman $w \in \mathcal{W}$. Indeed, the case where all the other women have popularity preferences is actually a linear combination of cases where those women have deterministic preferences. However, one needs to deal with the assumption that w receives at least one acceptable proposal during Algorithm 2.1.

Each person is either matched in all stable matchings or single in all stable matchings. For each person p, define X_p the event where p is matched. For every woman w, event X_w is true if and only if w is receives at least one acceptable proposal during Algorithm 2.1, which does not depend on the preference list of w. If we write Y_w the number of stable husbands of w, we have:

$$\forall w \in \mathcal{W}, \quad \mathbb{E}[Y_w \mid X_w] \leq 1 + \ln d_w + \mathbb{E}\left[\ln \frac{P_w(\mu_{\mathcal{W}}(w))}{P_w(\mu_{\mathcal{M}}(w))} \mid X_w\right]$$

We write $Y = \sum_{w \in \mathcal{W}} Y_w$ the total number of stable pairs. Using linearity of expectation we obtain:

$$\begin{aligned}\mathbb{E}[Y] = \sum_{w \in \mathcal{W}} \mathbb{E}[Y_w \mid X_w] \cdot \mathbb{P}[X_w] \leq & \sum_{w \in \mathcal{W}} (1 + \ln d_w) \cdot \mathbb{P}[X_w] \\ & + \sum_{w \in \mathcal{W}} \mathbb{E}[\ln(P_w(\mu_{\mathcal{W}}(w))) \mid X_w] \cdot \mathbb{P}[X_w] \qquad (5.6) \\ & - \sum_{w \in \mathcal{W}} \mathbb{E}[\ln(P_w(\mu_{\mathcal{M}}(w))) \mid X_w] \cdot \mathbb{P}[X_w] \qquad (5.7)\end{aligned}$$

The two sums in Equations (5.6) and (5.7) can be rewritten as sums over men: for every stable matching μ we have

$$\sum_{w \in \mathcal{W}} \mathbb{E}[\ln(P_w(\mu(w))) \mid X_w] \cdot \mathbb{P}[X_w] = \sum_{m \in \mathcal{M}} \mathbb{E}[\ln(P_{\mu(m)}(m)) \mid X_m] \cdot \mathbb{P}[X_m]$$

Using once again linearity of expectation:

$$\mathbb{E}[Y] \leq \sum_{w \in \mathcal{W}} (1 + \ln d_w) \cdot \mathbb{P}[X_w] + \sum_{m \in \mathcal{M}} \mathbb{E}\left[\ln \frac{P_{\mu_{\mathcal{W}}(m)}(m)}{P_{\mu_{\mathcal{M}}(m)}(m)} \mid X_m\right] \cdot \mathbb{P}[X_m]$$

To conclude the proof, observe that the sum $\sum_w \mathbb{P}[X_w]$ is at most $N = \min(M, W)$, and that the ratio $P_{\mu_{\mathcal{W}}(m)}(m)/P_{\mu_{\mathcal{M}}(m)}(m)$ is at most r_m. □

As long as the ratios r_m are polynomial in N, the expected number of stable pairs is $O(N \ln N)$.

5.5.2 Symmetric preferences

In this subsection, both men and women have popularity preferences. We say that popularities are *symmetric* when for every acceptable pair (m, w) we have $P_w(m) = P_m(w)$. To measure the extent to which popularities are symmetric, we introduce a parameter $r \geq 1$, such that for each acceptable pair (m, w) the values of $P_w(m)$ and $P_m(w)$ are within a factor r of each other.

Theorem 5.7. *Let $r \geq 1$ be a parameter. Assume that each person p has popularity preferences defined by P_p over a set of $d_p \geq 1$ acceptable partners, such that if man m and woman w are mutually acceptable then the values of $P_w(m)$ and $P_m(w)$ are within a factor r of each other. Then:*

$$\mathbb{E}[\textit{Number of stable pairs}] \leq N(1 + \ln r) + \sum_{w \in \mathcal{W}} \frac{\ln d_w}{2} + \sum_{m \in \mathcal{M}} \frac{\ln d_m}{2}$$

When $r = 1$, popularities are symmetric and $\mathbb{E}[\textit{Number of stable pairs}] \leq N + N \ln N$.

Proof. For each person p, define X_p the event where p is matched. We write Y the total number of stable pairs. Using the fact that women have popularity preferences, we start with the same proof as Theorem 5.6.

$$\mathbb{E}[Y] \leq \sum_{w \in \mathcal{W}} (1 + \ln d_w) \cdot \mathbb{P}[X_w] + \sum_{m \in \mathcal{M}} \mathbb{E}\left[\ln \frac{P_{\mu_{\mathcal{W}}(m)}(m)}{P_{\mu_{\mathcal{M}}(m)}(m)} \mid X_m\right] \cdot \mathbb{P}[X_m] \qquad (5.8)$$

Symmetrically, we can use the fact that men have popularity preferences, and use a symmetric version of Theorem 5.4 to bound the expected number of stable wife of each man.

$$\mathbb{E}[Y] \leq \sum_{m\in\mathcal{M}} \left(1 + \ln d_m + \mathbb{E}\left[\ln \frac{P_m(\mu_{\mathcal{M}}(m))}{P_m(\mu_{\mathcal{W}}(m))} \mid X_m\right]\right) \cdot \mathbb{P}[X_m] \tag{5.9}$$

Summing Equations (5.8) and (5.9) yields

$$\begin{aligned} 2\mathbb{E}[Y] \leq & \sum_{w\in\mathcal{W}} (1+\ln d_w)\cdot \mathbb{P}[X_w] + \sum_{m\in\mathcal{W}} (1+\ln d_m)\cdot \mathbb{P}[X_m] \\ & + \sum_{m\in\mathcal{M}} \mathbb{E}\left[\ln \frac{P_{\mu_{\mathcal{W}}(m)}(m)}{P_m(\mu_{\mathcal{W}}(m))} + \ln \frac{P_m(\mu_{\mathcal{M}}(m))}{P_{\mu_{\mathcal{M}}(m)}(m)} \mid X_m\right]\cdot \mathbb{P}[X_m] \end{aligned}$$

To conclude the proof, observe that the sums $\sum_w \mathbb{P}[X_w]$ and $\sum_m \mathbb{P}[X_m]$ are at most $N = \min(M, W)$, and that all ratios $P_w(m)/P_m(w)$ and $P_m(w)/P_w(m)$ can be bounded by r. □

5.6 High probability bounds

Recall that Theorem 5.4 gives an upper-bound on the expected number of stable partners of someone who has popularity preferences. The original upper-bound from Knuth, Motwanni and Pittel [KMP90] states that a person with uniform preferences has at most $(1+\varepsilon)\ln N$ stable partner, with probability $\to 1$ for every $\varepsilon > 0$ when $N \to +\infty$.

In Section 5.6.1 we give a high-probability bound on the number of stable partner of a woman having popularity preferences, when men who propose to her have bounded popularity, which holds in particular if she has bounded popularity preferences. In Section 5.6.2 we show that this last bound also applies when men have bounded popularity preferences and women have almost aligned popularity preferences.

5.6.1 Bounding the number of stable partners

The following theorem is a high probability analogue of Theorem 5.4.

Theorem 5.8. *Let w be a woman, we condition on $\mathcal{H}$ and we define the set of proposers as men who propose to her in Algorithm 2.2. Assume that w has popularity preferences defined by P_w over d_w acceptable partners, such that proposers are at most $C \geq 1$ times more popular than her partner $\mu_{\mathcal{M}}(w)$.*

$$\forall \varepsilon > 0, \quad \mathbb{P}[\textit{Nb of stable husbands of } w \geq (1+\varepsilon)\cdot(1+\ln d_w + \ln C)] \leq e^{-\frac{\varepsilon^2}{2+\varepsilon}(1+\ln C+\ln d_w)}$$

Proof. The proof continues the analysis started in Theorem 5.4. Conditioning on $\mathcal{H}$ and the preference of everyone except w, we know the sequence $x_1, \dots, x_K$ of men who propose to w in Algorithm 2.2. Let $Y = 1 + \sum_{i=1}^{K} X_i$ be the random variable equal to the number of stable partners of w, where each X_i is a Bernoulli random variable equal to 1 if man x_i is best so far. In the proof of Theorem 5.4, we used linearity of expectation to bound $\mathbb{E}[Y]$. Using Lemma 5.2, it turns out that X_i's are independent when w has popularity preferences, which is a useful property when using tail inequalities. Using Theorem 5.4, the expected value of Y is at most $1 + \ln d_w + \ln C$. Lemma 5.9 concludes the proof. □

Lemma 5.9 (Chernoff bound)**.** *Let* $S = \sum_{i=1}^{k} X_i$ *be the sum of* k *independent Bernoulli random variables, such that* $\mathbb{E}[S] \leq \mu$*. Then, for all* $\varepsilon > 0$ *we have* $\mathbb{P}[S \geq (1+\varepsilon)\cdot\mu] \leq e^{-\frac{\varepsilon^2\mu}{1+\varepsilon}}$*.*

Proof. The classical multiplicative Chernoff bound states that the Lemma holds if $\mu = \mathbb{E}[S]$. We artificially expend the sequence $X_1, \dots, X_k$ into a sequence $X_1, \dots, X'_k$ with $k' \geq k$, such that the expected value of $S' = \sum_{i=1}^{k'} X_i$ is equal to $\mathbb{E}[S'] = \mu$. Then we use the classical multiplicative Chernoff bound on S', and we use the fact that it stochastically dominates S. □

5.6.2 Bounding the popularity of proposers

When everyone has popularity preferences, the following theorem provides an upper bound on the popularity ratio between the stable husbands of a woman. The bound depends on how uniform the preferences of men are (parameter $R_{\mathcal{M}}$, the maximal ratio between the popularities of two distinct women for a given man) and how similar the preferences of women are (parameter $Q_{\mathcal{W}}$, the maximal ratio between the popularities of two distinct men for two distinct women). The parameter $R_{\mathcal{M}}$ is small when the preferences of every man among women are close to be uniform; they are actually uniform when $R_{\mathcal{M}} = 1$. The parameter $Q_{\mathcal{W}}$ is close to 1 when women tend to agree on the relative popularities of men. In case women have aligned popularity preferences, like in setting (1) of Theorem 5.1, then $Q_{\mathcal{W}} = 1$.

Theorem 5.10. *Assume that men and women have popularity preferences. Denote*

$$R_{\mathcal{M}} = \max_{\substack{m\in\mathcal{M}\\ w_0,w_1\in\mathcal{W}}} \frac{P_m(w_0)}{P_m(w_1)} \qquad Q_{\mathcal{W}} = \max_{\substack{w_0,w_1\in\mathcal{W}\\ m_0,m_1\in\mathcal{M}}} \frac{P_{w_0}(m_0)}{P_{w_0}(m_1)}\cdot\frac{P_{w_1}(m_1)}{P_{w_1}(m_0)} .$$

Let w *be a woman, we draw* $\mathcal{H}$ *and compute the maximum popularity* w *gives to each man who propose to her in Algorithm 2.2. With probability* $\geq (1-\frac{2}{N^2})$*, the ratio between this maximum popularity and the popularity of* $\mu_{\mathcal{M}}(w)$ *is no more than*

$$\left(N^5\cdot Q_{\mathcal{W}}\right)^{1+\frac{4\ln(N)(1+\log_2(N))}{\ln(1+1/R_{\mathcal{M}})}} .$$

Proof of Theorem 5.10. Fix a woman w, and set

$$T = |N|^5.$$

We define the notion of *standard preferences* for women $\neq w$. Let m, m' be two men and w' a woman such that m' is T times more popular than m but, still, w' prefers m to m'. We say that the preferences of women $\neq w$ are *standard* if no such tuple m, m', w' exists. Given m, m', there is probability $\frac{1}{1+T}$ that w' prefers m to m', thus there is probability $\leq \frac{1}{N^2}$ that preferences of women are not standard. In the rest of the proof we draw the preferences of all women except w assume that those preferences of are standard.

We now draw $\mathcal{H}$ and denote $\mu_{\mathcal{M}}$ be the man-optimal stable matching. In the sequel, we reference as "w-popularities" the popularities from the point of view of w given by P_w. The proof relies on the computation of the set of stable husbands of w by Algorithm 2.2. The algorithm is presented as running deterministically, with all preferences of men and women given as input, chosen *ex-ante* before the algorithm starts; however in the case of popularity preferences (Definition 3.2) this algorithm can also be seen as a stochastic process.

In phase 1 of Algorithm 2.2, instead of computing the lists of men *ex-ante*, we might instead disclose them progressively along the execution of the algorithm. When a man is about to propose, he randomly picks a woman among those to whom he has not proposed yet, following a lottery whose probabilities are proportional to the popularities of the remaining women. When a woman $\neq w$ receives a proposal, her answer is deterministic, consistent with her preferences fixed ex-ante. When w receives a proposal, she refuses it.

In this stochastic process, we focus on the sequence $(x_t)_{t\geq 0}$ of men that are enumerated by Algorithm 2.2, after the initial computation of the man-proposing stable matching $\mu_{\mathcal{M}}$. The sequence includes all men doing a proposal, including proposals to w but also to other women. A man appears in the list as many times as he makes a proposal. The set of stable husbands of w will be exactly the set of men from whom w accepts such a proposal.

Without loss of generality, we assume that men $m_1, m_2, \ldots, m_M$ are indexed by increasing w-popularity, and let i be the index of the husband of w in the man-proposing stable matching $\mu_{\mathcal{M}}(w) = m_i$.

$$P_w(m_1) \leq P_w(m_2) \leq \ldots \leq P_w(m_N).$$

We partition the set of men in sets of exponentially increasing sizes, starting with all men less or equally popular than m. Let $F_0 = [1, i]$, $F_1 = (i, 2i]$, $F_2 = (2i, 4i]$, ..., $F_j = (2^j \cdot i, N]$, where $2^j \cdot i < N \leq 2^{j+1} \cdot i$. Note that $j \leq \ln_2(N)$. Set

$$L = \frac{4\ln(N)}{\ln\left(1 + \frac{1}{R_{\mathcal{M}}}\right)} \qquad K = (T \cdot Q_{\mathcal{W}})^L.$$

For every set F_ℓ with $1 \leq \ell \leq j$, we say that *there is a huge popularity gap* in F_ℓ if the popularity ratio for w in this interval is $\geq K$, i.e. if $KP_w(m_{2^\ell \cdot i}) \leq P_w(m_{\min(N, 2^{\ell+1} \cdot i)})$.

The case where there is no huge popularity gap in any of the F_ℓ with $1 \leq \ell \leq j$ is an easy one: then the most w-popular man m_N has w-popularity at most $K^{j+1} P_w(m_i)$ hence the conclusion of the theorem since $j \leq \ln_2(N)$.

Otherwise we select the smallest ℓ for which there is huge popularity gap in F_ℓ. Denote $E_0 = F_0 \cup F_1 \cup \ldots F_{\ell-1}$ and $E_1 = F_\ell$. Define to popularity thresholds $p_{inf} = P_w(m_{2^\ell \cdot i}) \cdot (T \cdot Q_{\mathcal{W}})$ and $p_{sup} = P_w(m_{2^{\ell+1} \cdot i})/(T \cdot Q_{\mathcal{W}})$, and decompose $E_1 = E_1^{bot} \cup E_1^{top} \cup E_1^{top}$ such that

$$\begin{aligned} E_1^{bot} &= \{m \in E_1 \mid P_w(m) < p_{inf}\} \\ E_1^{mid} &= \{m \in E_1 \mid p_{inf} \leq P_w(m) < p_{sup}\} \\ E_1^{top} &= \{m \in E_1 \mid p_{sup} \leq P_w(m)\} \end{aligned}$$

Observe that the sequence $(x_t)_{t\geq 0}$ satisfies the following conditions:

i) The w-popularities of two consecutive men in the sequence $(x_t)_{t\geq 0}$ may increase by a multiplicative factor of at most $T \cdot Q_{\mathcal{W}}$.

ii) If $x_t \in E_0$, then $x_{t+1} \in E_0 \cup E_1^{bot}$. If $x_t \in E_1^{bot}$, then $x_{t+1} \in E_0 \cup E_1^{bot} \cup E_1^{mid}$.

iii) Assume that in the current execution of the algorithm, the proposing man is $x_t \in E_1^{mid}$. Then $x_{t+1} \in E_0 \cup E_1$, and the probability that x_{t+1} belongs to E_0 conditioned on the current execution is $\geq \frac{1}{R_{\mathcal{M}}+1}$.

iv) Assume that $x_t \in E_1^{top}$ for the first time. Then the sequence stayed in E_1^{mid} the last L consecutive steps. The probability of this happening at any time $t \leq N^2$ is at most $\frac{1}{N^2}$.

Properties i) - iii) follow from the hypothesis that women $\neq w$ have standard preferences: for a woman $w' \neq w$ married to x_{t+1} to accept a proposal from x_t it is necessary (but not sufficient in general) that $\frac{P_{w'}(x_{t+1})}{P_{w'}(x_t)} < T$. In which case $\frac{P_w(x_{t+1})}{P_w(x_t)} < T \cdot Q_{\mathcal{W}}$. And if $w' = w$ then w will anyway refuse the proposal from x_t thus $x_{t+1} = x_t$. Hence i) and ii). To prove iii), remark that whenever some man x_t from E_1^{mid} proposes to the wife of a man m, the wife will refuse (resp. accept) for sure if m is outside $E_0 \cup E_1$ (resp. is in E_0), because in that case m is at least $T \cdot Q_{\mathcal{W}}$ times more w-popular (resp. less w-popular) than x_t. Since E_0 contains at least half of the men in $E_0 \cup E_1$ and since women married to men in E_0 are at most $R_{\mathcal{M}}$ times less popular for x_t than women married to men in E_1, we get iii). To prove iv), observe men in E_1^{top} are at least $(Q_{\mathcal{W}} \cdot T)^L$ times more popular than men in E_1^{bot}, hence the sequence must stay in E_1^{mid} at least $L - 1$ consecutive steps before reaching E_1^{top}. Applying iii) repeatedly shows that for every t the stochastic process $x_t, x_{t+1}, \ldots$ reaches E_1^{top} with probability at most

$$\left(1 - \frac{1}{R_{\mathcal{M}}+1}\right)^L = \exp\left(-\ln\left(1 + \frac{1}{R_{\mathcal{M}}}\right) \cdot L\right) = \frac{1}{|N|^4}.$$

Since the length of the sequence $(x_t)_{t\geq 0}$ is bounded by N^2, we get iv) by union bound. Thus, the probability that a proposer is K times more popular than m_i is smaller than $1/N^2$. Combining this with the probability that preferences of women are standard concludes the proof. □

5.7 Simulations

Implementations of two-sided matching markets with popularity preferences are shared between Chapters 4, 8 and 5, and are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/Popularity

Recall that Figure 5.1 illustrates the effect of correlations on the number of stable pairs: panel (a) has positively correlated preferences and few stable pairs, panel (b) has uncorrelated preferences and some stable pairs, and panel (c) has negatively correlated preferences and many stable pairs. In Figures 5.3, 5.4 and 5.5, we go smoothly from the instance from panel (b) to the instance from panel (c), replacing uniform preferences by cyclic preferences. We show that the number of stable pairs increases when introducing negative correlations.

Notice that in the simulations of Chapter 4, we go smoothly from the instance from panel (a) to the instance from panel (b), assuming that women have popularity preferences $P : m_i \mapsto \lambda_{\mathcal{M}}^i$ with $\lambda_{\mathcal{M}} = 0.5$, and that men have popularity preferences $P : w_i \mapsto \lambda_{\mathcal{W}}^i$ with $\lambda_{\mathcal{W}} \in \{0.5, 0.9, 1\}$. Interestingly, we show that the number of agents with multiple stable partners is minimal when $\lambda_{\mathcal{W}} = 1$, which might seems counter intuitive given our claim that "positive correlations decrease the number of stable pairs". This can be explained by the fact that having $\lambda_{\mathcal{W}} = 1$ unbalances the market and creates a nearly unique stable matching. This shows that looking at correlations is not the end of the story!

5.8 Conclusion and open questions

In this chapter, we discuss "weak core-convergence", in terms of number of stable pairs. We show that uniformly random preferences is a worst case when compared to positively correlated preferences. The following questions are left open for future work:

- **Number of stable partners.** In Theorem 5.1, we bound the total number of stable pairs, but not individual numbers of stable partners. When both sides have aligned popularity preferences, we conjecture that each person has at most $1 + \ln N$ stable partners in average. Notice that this is true when one side has uniform preferences, because of the additional symmetry.

- **Removing cross-sided correlations.** In the introduction, we argue that negative correlation can increase the number of stable pairs. In particular, we believe that if the market has no negative cross-sided correlations (people like people who do not like them in return), then the total number of stable pairs is at most the one from the uncorrelated case. Formally speaking, we can remove cross-sided correlations from an adversarial instance by shuffling the preference list of agents on each side of the market. We conjecture that for any instance, the expected total number of stable pairs after shuffling is at most $N(1 + \ln N)$. In particular, the conjecture appears to be true on the instance from Figures 5.2 and 5.5. The conjecture is false if we only shuffle preferences lists of agents from one side.

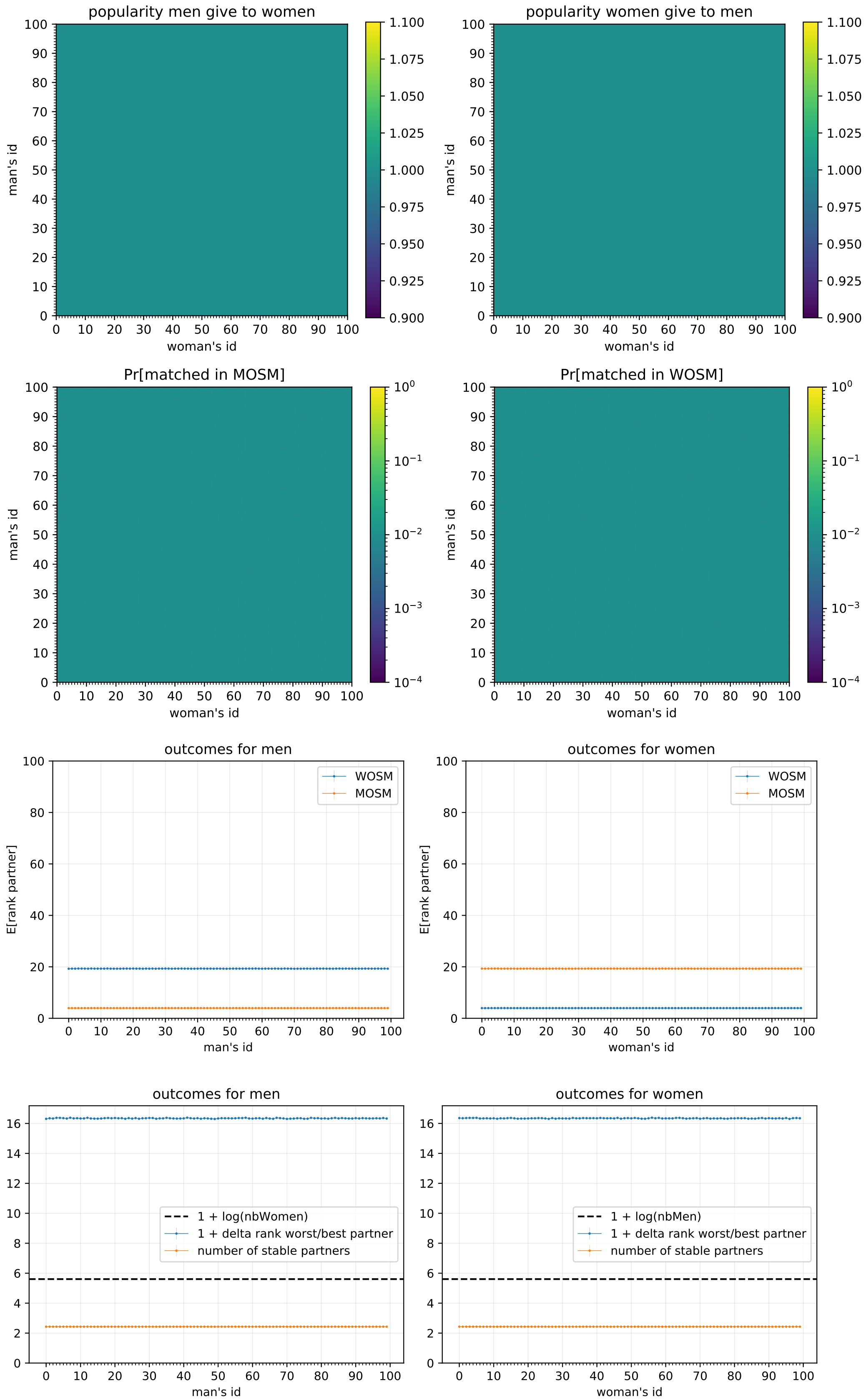


Figure 5.3. Matching market with $N = 100$ men and $N = 100$ women having uncorrelated preferences. From [Pit89; KMP90; Pit92], in expectation each person has $\sim \ln N$ stable partners, ranked between $\sim \ln N$ for the best and $\sim N/\ln N$ for the worst.

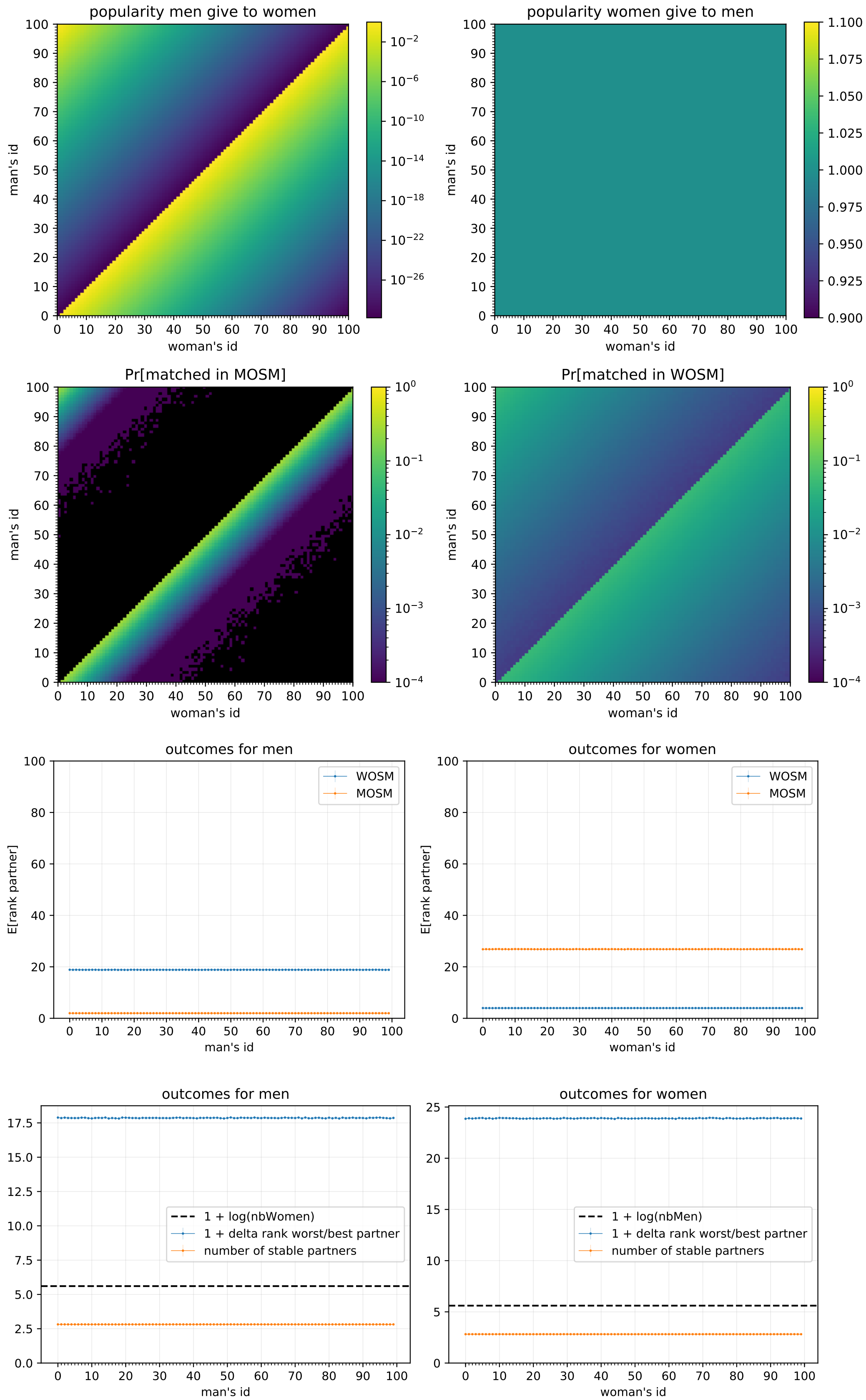


Figure 5.4. Matching market with $N = 100$ men and $N = 100$ women, where women have uncorrelated preferences and men have negatively correlated preferences. Both the number of stable partners and the rank difference between the worst and best partners are bigger than in Figure 5.3.

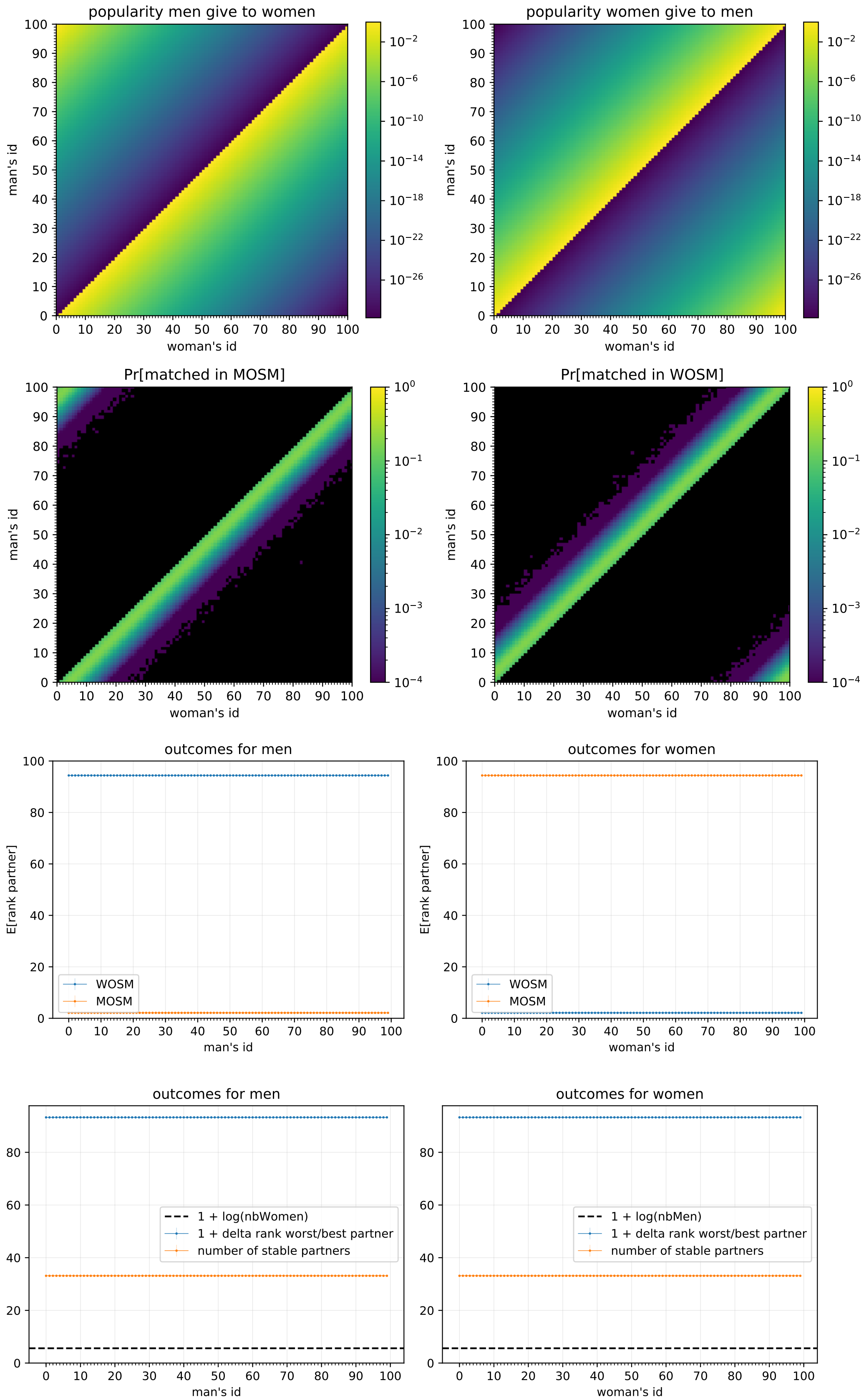


Figure 5.5. Matching market with $N = 100$ men and $N = 100$ women, where men and have negatively correlated preferences, both within a side (preferences are cyclic) and across sides (preferences are reversed). Both the number of stable partners and the rank difference between the worst and best partners are bigger than in Figure 5.4.

6 | Constrained School Choice with Incomplete Information

This chapter is based on the following paper:

[GMM21a] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Constrained School Choice with Incomplete Information". In: *arXiv preprint arXiv:2109.09089* (2021)

6.1 Introduction

School choice is referred in the literature as the two-sided matching market where students (on one side) are to be matched with schools (on the other side) based on their mutual preferences. As discussed in previous chapters, a classical solution concept is the celebrated deferred acceptance procedure, proposed by Gale and Shapley [GS62], and since implemented by many clearinghouse [AS03; APR05; Cor+19]. Most often in practice, the clearinghouse sets an upper quota on the number of applications each student can submit. This requires a strategic behaviour from students who should find a balance between applications to top-tier schools and applications to less attractive but also less selective lower-tier schools.

The existence and computability of Nash equilibria is a desirable property for two reasons. First, Nash equilibria are among the possible long-term outcomes of the market, possibly emerging after a series of best-response dynamics or evolutional selection of strategies. Second, and most importantly, being able to compute a Nash equilibrium provides a solid basis to develop a recommendation system in order to help the students to select the schools they want to apply to.

In case complete information about the preferences of the students and schools is available, implementing a Nash equilibrium is rather easy. The strategic behaviour of students limited to a fixed number of options was studied by Romero-Medina [Rom98], and later investigated by Calsamiglia, Haeringer and Klijn [HK09; CHK10]. After a pre-computation of the student-optimal stable matching $\mu_{\mathcal{S}}$, a simple recommendation can be made to every student matched in $\mu_{\mathcal{S}}$: they only need to apply to a single school, their match in $\mu_{\mathcal{S}}$. As a direct corollary of [DF81; Rot82], this leads to a Nash equilibrium. Remark that student unmatched in $\mu_{\mathcal{S}}$ wont be matched in this equilibrium, whatever strategy they choose.

In practice, assuming that the student-optimal stable matching is computable *ex-ante* is rather unrealistic. For example, in the French college admission system "Parcoursup", there are more than 900000 students signed up. Applicants should report a shortlist of 20 wishes before a fixed deadline. Based on statistics of previous years, one might evaluate how the grades of a particular student compare to others and evaluate the percentage of students that will be ranked higher in a particular school. But acquiring before the deadline the information needed for an exact computation of the student-optimal matching is unfeasible.

In this chapter, we propose a formal model for the constrained school choice with incomplete information, and study the existence and computability of Nash equilibria in the associated incomplete information game. In our model, each student draws a type from a publicly known distribution μ (see Section 6.2). In Section 6.3, we detail interesting examples that can be used to state recommendations for students, schools and decision makers. In Section 6.4, we give the proof of existence of a symmetric Bayes-Nash equilibrium (Theorem 6.4). In Section 6.5 we give efficient algorithms to compute equilibria when the number of types is finite and additional hypotheses are made, including the case where schools have identical preferences over students (Theorems 6.6, 6.9 and 6.8). In Section 6.6 we prove a convergence theorem, showing that one can compute an equilibrium for a game with a continuous type distribution μ, using a (weakly) converging sequence of distributions $(\mu_k)_{k\geq 1}$ having finite supports (Theorem 6.12).

Related work. This chapter is closely related to the literature of matching under random preferences. Pittel [Pit92] study balanced matching markets with uniformly random preferences. Rephrasing his results in our setting, the student-proposing deferred acceptance procedure matches with high probability every student to one of her top $\log^2 n$ choices, which proves that an upper quota of $\log^2 n$ applications per student does not deteriorate the outcome. Immorlica and Mahdian [IM15], and Kojima and Pathak [KP09] study matching markets where one side of the market has random preference lists of constant size, and show that such markets have a (nearly) unique stable matching. When quotas are constant and preferences are uniform, this implies that games based on student or school proposing deferred acceptance are (almost) the same.

More recent papers discuss the effect of an upper quota on the number of applications. Beyhaghi, Saban and Tardos [BST17] study the efficiency of equilibria in a model where each side is divided into two uniform tiers, and each student chooses her number of applications to top-tier schools. Beyhaghi and Tardos [BT21] study the social welfare (size of the matching) as a function of the number of applications, in a model where preferences of agents are drawn uniformly at random. Echenique, Gonzalez, Wilson and Yariv [Ech+20] examine the National Resident Matching Program and argue that doctors are strategic when reporting their preferences.

The best response of a student to the strategies of others is related to the simultaneous search literature. Chade and Smith [CS06] discuss the problem where one student must choose a portfolio of schools in which she applies: each application has a cost, a probability of success and a cardinal utility when successful. Ali and Shorrer [AS21] generalize their model to allow correlations between admission decisions.

Takeaway message. In general, the deferred acceptance mechanism is known to be strategy-proof for the proposing side [DF81], but no mechanism is truthful for both sides of the market [Rot82]. However, empirical results show that the stable matching is often unique [RP99], in which case stable matching procedures are truthful for all agents, even when they have incomplete information [EM07]. Thus, having a unique stable matching is a desirable property, and we argue that this fact carries over to the case where students have restricted preferences. First, in terms of number of equilibria, examples (see Section 6.3) illustrate that the fewer stable matchings there are, the fewer equilibria the game has. Second, in terms of outcome, multiple stable matchings can induce outcomes which are unstable (see Section 6.3.1) or sub-optimal (see Section 6.3.2). And finally, in terms of computability of an equilibrium, Section 6.5 give two algorithms to compute equilibria in time quasi-linear w.r.t. the number of types, under the extra assumption that markets induced by the game have unique stable matchings.

6.2 The model

We consider a game where players are students who do not know the exact preferences of other students. For the sake of modeling, each student has a type $T = [0,1]^d$ with $d \geq 1$, which can be thought as a feature vector representing both her preferences and characteristics. Types are drawn without replacement[1] from the set of types, using a probability distribution $\mu \in \Delta(T)$. Each student i knows her own type $t_i \in T$ (private information) and the distribution μ (common information). Each school j has a capacity c_j, a bounded measurable value function $v_j : T \to \mathbb{R}_+$ and a measurable scoring function $s_j : T \to [0,1]$.

The set of actions A is the set of preference lists containing at most ℓ schools[2]. Each student i reports a preference list $a_i \in A$. Schools sort students by decreasing score, breaking ties uniformly at random. Then, we compute a matching using the student proposing deferred acceptance algorithm. Each student i receives a utility $v_j(t_i)$ if she is assigned to school j, and a utility of 0 is she stays unmatched.

Algorithm 6.1 Description of the matching game

Game parameters: n, m, A, T, $(v_j)_{j\in[m]}$, $(c_j)_{j\in[m]}$ and $(s_j)_{j\in[m]}$.
Function $\textsc{Utility}((t_i, a_i)_{i\in[n]} \in (T \times A)^n)$
  Student i has type t_i, reports the preference list a_i, and her score at school j is $s_j(t_i)$.
  School j has capacity c_j, sorts students by decreasing scores,
    breaking ties uniformly at random.
  Students are assigned to schools using the student proposing deferred acceptance algorithm.
  Each student i receives utility $u_i = v_j(t_i)$ if she assigned to school j,
    and $u_i = 0$ if she is unassigned.
  Return the vector of utilities $(u_i)_{i\in[n]}$, averaged over all possible tie-breaking choices.

Students choose their actions strategically: the set of (behavioral) strategies $\mathcal{S}$ is the set of measurable function $p : T \to \Delta(A)$. A strategy profile is a vector of strategies $(p_i)_{i\in[n]} \in \mathcal{S}^n$, where p_i is the strategy of student i and p_{-i} denotes the vector of strategies of all students except i. Under this strategy profile, the expected payoff of student i is denoted $U_\mu(p_i, p_{-i})$, this is the i-th component of the vector $\mathbb{E}_{t,a}[\textsc{Utility}((t_i, a_i)_{i\in[n]})]$, where the expectation is taken over the random draws of t and a: t_i's are drawn without replacement from μ and each a_i is drawn from $p_i(t_i)$. Notice that this definition already incorporates the symmetry of the game: if the strategies in p_{-i} are permuted, the payoff of player i does not change, and if we swap two players, their payoffs are swapped accordingly. In other words, the payoff of a player only depends on his own strategy and the multiset of strategies played by the other players, independently of each player's identity.

A strategy profile is a Bayes-Nash equilibrium if each student cannot improve her utility by deviating from the strategy profile. More precisely, $(p_i)_{i\in[n]} \in \mathcal{S}^n$ is an equilibrium if $U_\mu(p^*, p_{-i}) \leq U_\mu(p_i, p_{-i})$ for every $i \in [n]$ and $p^* \in \mathcal{S}$.

Our first theorem states that the strategic part of the game for a student is to choose her (unordered) set of applications. More precisely, once she decided which schools she will apply to, it is optimal for her to sort schools by decreasing value. As a Corollary, when $\ell = m$, the set of actions A is unconstrained and contains all the permutations over schools, thus sorting schools by decreasing score is a dominant strategy.

---

[1]Types are drawn without replacement in order to have a well defined game when the distribution μ is discrete. This does not mean students cannot have the same preferences over schools, as one can duplicate types by increasing the dimension d of the type space. When the distribution is non-atomic, it is equivalent with having types drawn independently.

[2]More generally, results of this paper hold if A is an arbitrary subset of preference lists over schools.

Theorem 6.1. *Let $t \in T$ be a type and $a \in A$ be an action, and define a^* the preference list where schools from a are sorted by non-increasing order of value $v_j(t)$. If a^* is a valid action, then for a student of type t reporting a^* dominates reporting a.*

Proof. Student proposing deferred acceptance is truthful for students (see Corollary 2.24). □

6.3 Motivating examples

6.3.1 Complete information

Recall that types of students are drawn without replacement from μ. Thus, if μ is a discrete distribution with a finite support of size n, then students exactly know the types of other students, which proves that complete information is a special case of our model.

Haeringer and Klijn [HK09] study an equivalent complete information game: n students and m schools have ordinal preferences over one another, and each student must report preference lists of length at most ℓ to the clearinghouse. When $\ell = 1$, they show that each stable matching can be implemented at equilibrium (Proposition 6.1), and the outcome of every equilibrium is a stable matching (Proposition 6.3). Additionally[3], they give examples to show that when $\ell > 1$, the outcome of some equilibrium can be unstable (Examples 6.6, 8.3, 8.4 and 8.5). In Figure 6.1, we reproduce Example 8.3 from [HK09].

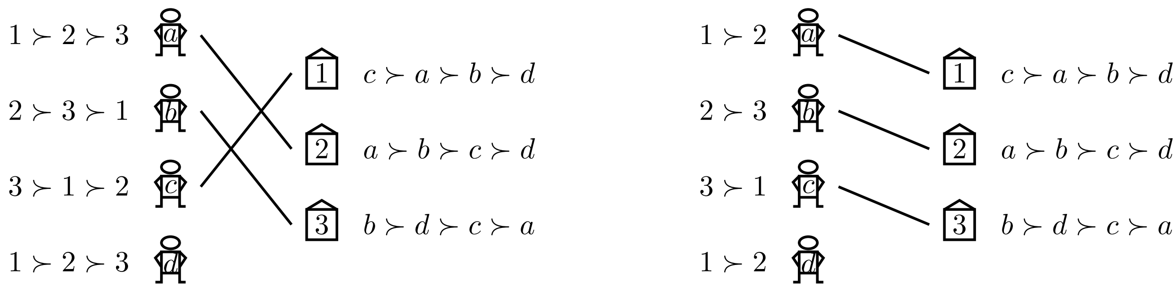


Figure 6.1. The true preferences of students are displayed on the left: there is a unique stable matching where students a, b and c get their 2nd choices. When students are restricted to apply to at most two schools, truncating each student preference list yields an equilibrium: students a, b and c get their 1st choice, whereas d will stay unmatched no matter what she reports. The outcome is not stable as student d and school 3 prefer each other to their respective partners. Observe that if d applies to school 3 then students a, b and c will get their 2nd choices, which is a different equilibrium where the outcome is stable.

In general, because every stable matching can be implemented by an equilibrium, a necessary condition for the outcome to be unique is to have a unique stable matching. This condition however is not sufficient, as Figure 6.1 illustrates with an example having a unique stable matching but several possible outcomes when $\ell = 2$. In Theorem 6.3, we show that α-reducibility is a sufficient condition for having a unique outcome. The notion of α-reducibility was introduced by Alcade [Alc94] in the context of stable roommates, then investigated by Clark [Cla06] who showed it is equivalent to having a unique stable matching in every sub-market (Theorems 4 and 5).

[3]When $\ell \geq 1$, Haeringer and Klijn [HK09] also give (Theorem 6.6) a necessary and sufficient condition on the preferences of schools such that the outcome is stable for every preferences of students and for every equilibrium. This result is in general incomparable with our Theorem 6.3

Definition 6.2 (α-reducibility)**.** We say that a two-sided matching market is α-reducible if for every subset of students $A \subseteq [n]$ and subset of schools $B \subseteq [m]$, there exist a fixed pair $(i, j) \in A \times B$ such that i and j strictly prefer each other to everyone else in A and B.

Theorem 6.3. *If the two-sided matching market is α-reducible, then for every $\ell \geq 1$ the outcome of every Nash equilibrium is the unique stable matching.*

Proof. For every Nash equilibrium, start the analysis by setting $A = [n]$ and $B = [m]$. From α-reducibility we know that there is a fixed pair $(i, j) \in A \times B$. Student i can ensure she is matched with her first choice j, thus this must be her outcome by definition of a Nash equilibrium. We remove i from A, decrease the capacity of j and remove it from B if it reached 0. We continue with the same reasoning by induction. □

6.3.2 Reversed preferences

Consider a simple example with $n = 2$ students and $m = 2$ schools of capacity $c_1 = c_2 = 1$. The set of types $T = [0, 1]^2$ is two-dimensional. A student of type (x, y) gives the value $v_1(x, y) = r + x$ to school 1, and the value $v_2(x, y) = r + y$ to school 2, where r is a positive constant representing how risk-averse the students are. The preferences of schools are reversed, in the sense that a student of type (x, y) has a score of $s_1(x, y) = y$ at school 1 and a score of $s_2(x, y) = x$ at school 2. Figure 6.2 illustrates a situation with two stable matchings. This occurs with probability 1/2 when μ is uniform over the diagonal $\{(x, y) \in T \mid x + y = 1\}$, and with probability 1/6 when the distribution μ is uniform over T.

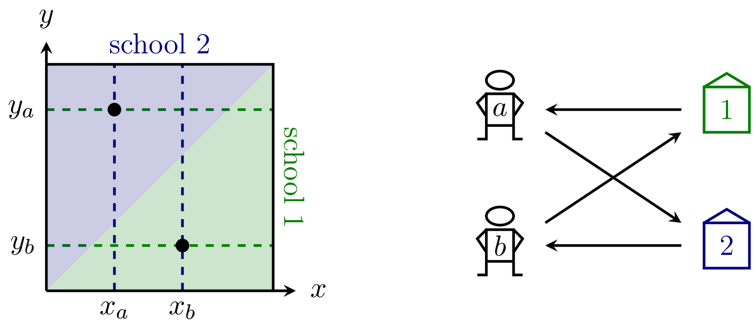
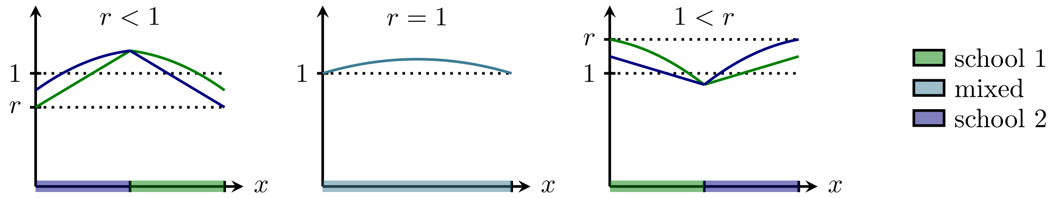
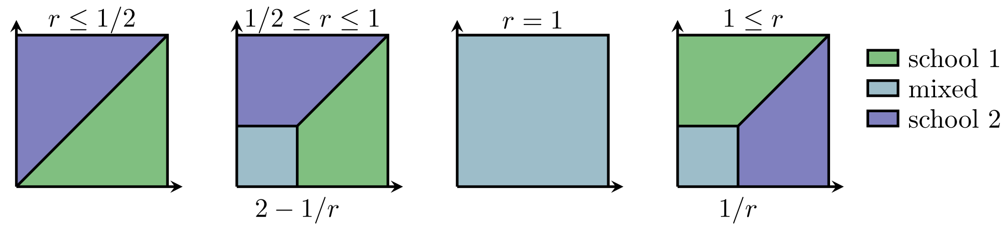


Figure 6.2. An example with two stable matchings: student a prefers school 2 (upper triangle $x_a < y_a$), student b prefers school 1 (lower triangle $y_b < x_b$), school 1 prefers student a (horizontal lines $y_b < y_a$), and school 2 prefers student b (vertical lines $x_a < x_b$).

Risk aversion. In Figure 6.2, if students were allowed to apply to both schools, the student proposing deferred acceptance procedure would always choose the student optimal stable matching. However, for some reason, the clearinghouse only allows students to apply to one school. In order to maximize their expected utilities, students can either prioritize the value they give to schools, or the likelihood of being accepted. Figure 6.3 illustrates two families of Bayes-Nash equilibria, as a function of r: the more risk averse students are, the less likely is the student optimal matching to be chosen.

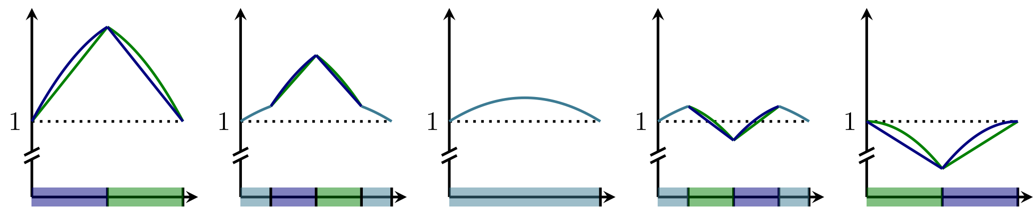


(a) Family of equilibria when the distribution μ is uniform over the diagonal $\{(x, y) \in T \mid x + y = 1\}$. The strategy of a student having type $(x, 1 - x)$ is represented by the color of the point on the x-axis. The different plots corresponds to the expected utility from each action, for a student of type $(x, 1-x)$, when the other students play the equilibrium strategy.

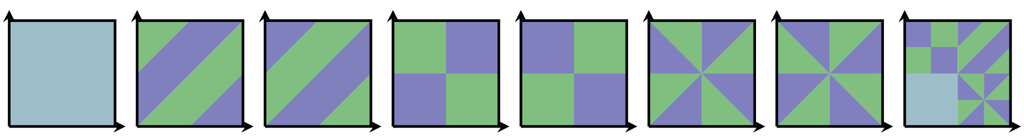


(b) Family of equilibria when the distribution μ is uniform over T. The strategy of a student having type (x, y) is represented by the color of the point at those coordinates.

Figure 6.3. Family of symmetric Bayes-Nash equilibria when $\ell = 1$ in the game described in Figure 6.2. When students are not risk averse ($r < 1$), they tend to apply according to their preferences. When students are risk averse ($r > 1$), they tend to apply according to their chance of being accepted.



(a) Five different equilibria when the distribution μ is uniform over the diagonal $\{(x, y) \in T \mid x+y = 1\}$. Students' utility induce a total ordering over the family of equilibria, where the leftmost equilibrium is student optimal and the rightmost equilibrium is student pessimal.



(b) Eight different equilibria when the distribution μ is uniform over T. Notice that for every fixed x (or y), the fractions of types applying to schools 1 and 2 is the same. Thus, in each equilibrium, a student of type (x, y) is accepted to school 1 (resp. 2) with probability $(1 + y)/2$ (resp. $(1 + x)/2$) and her expected utility is $(1 + x)(1 + y)/2$ for both schools.

Figure 6.4. Multiple Bayes-Nash equilibria when $\ell = 1$ and $r = 1$ in the game described in Figure 6.2. Such phenomenon can be explained by the multiplicity of stable matchings (see panel a), or by a "purification theorem" type of argument (see panel b).

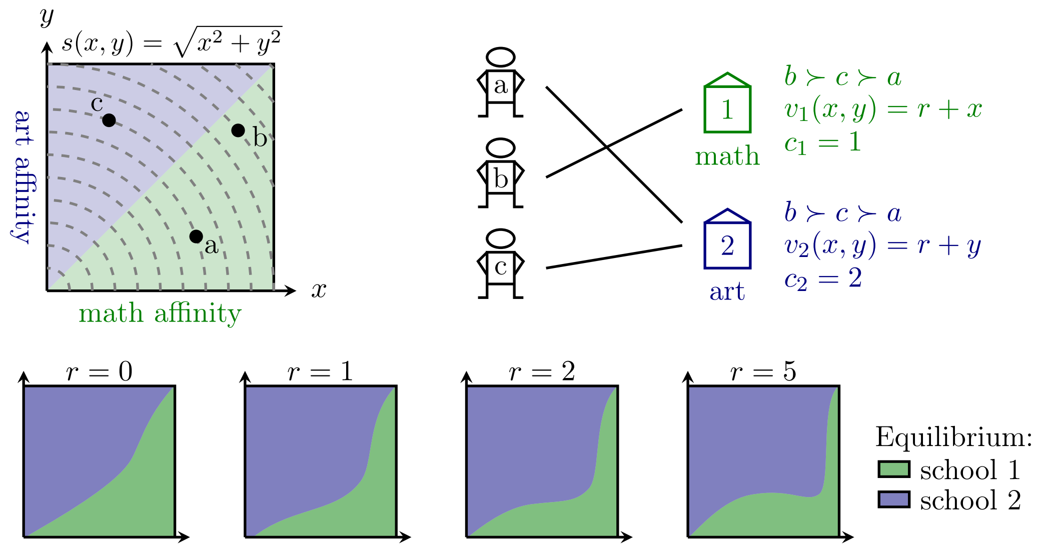


Figure 6.5. Example where preferences of schools are aligned (scoring functions are equal to s). The two dimensions of a student's type can be thought as her affinity with maths and with art. Schools sort student by decreasing order of euclidean norm. School 1 is a math school, and is preferred by students with a good affinity with math. School 2 is an art school, and is preferred by students with a good affinity with art. Risk aversion is modeled with the parameter $r \geq 0$. Each student can apply to $\ell = 1$ school, and the strategy of a student having type (x, y) is represented by the color of the point at those coordinates.

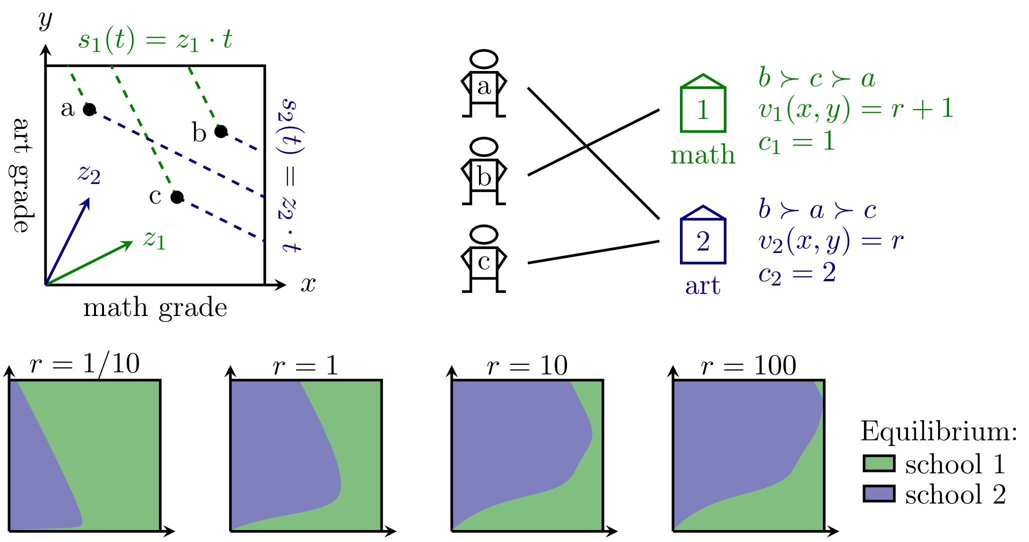
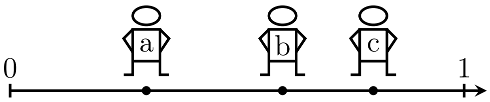


Figure 6.6. Example where preferences of students are aligned (value functions are constant). The two dimensions of a student's type can be thought as her grades in maths and in art. School 1 has one seat and gives more importance to the math grade. School 2 has two seats and gives more importance to the art grade. Risk aversion is modeled with the parameter $r \geq 0$. Each student can apply to $\ell = 1$ school, and the strategy of a student having type (x, y) is represented by the color of the point at those coordinates.

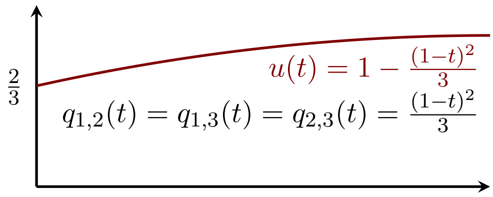
$n = 3$ students, μ is uniform over $T = [0, 1]$



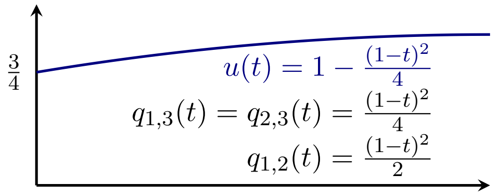
$m = 3$ identical schools:

- capacities $c_1 = c_2 = c_3 = 1$
- values $v_1 = v_2 = v_3 = 1$
- scores $s_1(t) = s_2(t) = s_3(t) = t$

Denote $q_{i,j}(t)$ the probability that a student of type t will be rejected from schools i and j, because the other two students have types $> t$ and have already been assigned to those two seats. Denote $u(t)$ the expected utility of a student of type t at equilibrium.



Equilibrium #1: every student draws a list uniformly at random, among the 6 possible lists of length 2.



Equilibrium #2: every student chooses uniformly at random between the lists $(1, 3)$ and $(2, 3)$.

Figure 6.7. Example with $n = 3$ students, $m = 3$ identical schools, and $\ell = 2$ applications per student. If students had perfect information or where allowed to apply to all 3 schools, they will all be assigned and will always receive a payoff of 1. When students have imperfect information and can only apply to 2 schools, every symmetric equilibrium will leave a student unassigned with positive probability. The intuition behind the improved payoff in equilibrium #2 is that students privately agree that school 3 is a "safety choice".

$n = 50$ students, μ is uniform over $T = [0, 1]$
$m = 5$ schools, with identical preferences $s(t) = t$

1. value $v_1 = 5$, capacity $c_1 = 5$. → students of rank 1 to 5
2. value $v_2 = 4$, capacity $c_2 = 10$. → students of rank 6 to 15
3. value $v_3 = 3$, capacity $c_3 = 5$. → students of rank 16 to 20
4. value $v_4 = 2$, capacity $c_4 = 10$. → students of rank 21 to 30
5. value $v_5 = 1$, capacity $c_5 = 10$. → students of rank 30 to 40

R_t = rank of a type t student
$\mathbb{E}[R_t] = 1 + (n-1)(1-t)$
$\mathbb{P}[R_t \le r] = I_t(n - r, r)$

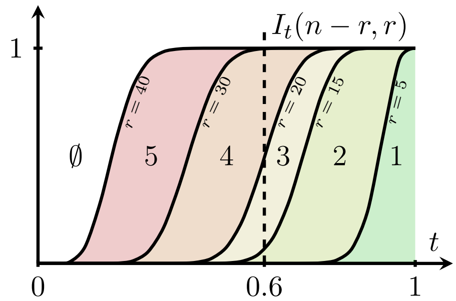


Figure 6.8. Example where preferences of schools and students are aligned. If students are allowed to apply to all the schools, their outcome is determined by their rank. Let R_t be the rank of a student of type t. Then $R_t - 1$ follows a binomial distribution of parameter $(n - 1, 1 - t)$, whose cumulative function can be expressed using the regularized incomplete beta function I. In particular, a student of type $t = 0.6$ will be assigned to school 3 with probability $I_{0.6}(30, 20) - I_{0.6}(35, 15) \approx 43\%$. For every t, one could observe that the combined probability of the 2 least likely outcomes never exceeds 2%. Hence when students are restricted to 3 applications, each student can, at least 98% of the time, ensure the same outcome as the one when everyone has 5 applications.

Multiple equilibria. In the complete information case, every stable matching induces an equilibrium. In the incomplete information case, having multiple stable matchings can induce an infinite number of equilibria. Figure 6.4(a) gives an example with a continuum of equilibria, such that the expected payoff of each student type is non-increasing (from left to right). When multiple stable matchings exist, the left-most equilibrium always implement the student optimal stable matching, and the right-most equilibrium always implement the school optimal stable matching. Conversely, Figure 6.4(b) gives an example with an infinite number of equilibria where each student type receives exactly the same expected utility from every equilibrium.

6.3.3 Aligned preferences

Computing an equilibrium by induction. When preferences of schools are identical (scoring functions are equal), the best student can ensure she will be matched with her favorite school. When preferences of students are identical (value functions are constant), the student ranked first by the best school can ensure she will be matched with her favorite school. An equilibrium for such games can be computed by eliminating dominant strategies: when each type's payoff does not depend on the strategies of types having lower scores, we can proceed by induction. Figures 6.5 and 6.6 give two examples of such games, and illustrate the type of recommendation one could provide to students using this approach.

Identical schools should merge their selection process. Theorem 6.3 shows that if the matching market is α-reducible and students have perfect information, then every equilibrium yields the same outcome. This is true in particular when all schools have identical preferences, independently of the preferences reported by students. It is a natural question to ask if the same hold with incomplete information. Figure 6.7 provides a counter-example with three students and three identical schools. If students can only apply to two schools, one student will stay unassigned with positive probability. One way for the students to reduce this probability is to privately agree that one of the school is a "safety choice" which is always ranked last. Even if schools were *a priori* identical, such strategies impact the quality of students selected in the safety school, which may cause a differentiation between the schools from one year to the next. Such an example could be interpreted as a recommendation for identical schools to merge their selection process.

Number of applications. Figure 6.8 gives an example where both the preferences of students and schools are aligned. In such case there is a unique stable matching where the outcome of a student is determined by her rank: the best students are matched to the best school, the next students are matched to the second school, and so on. A student who has incomplete information and only knows her expected rank can try to apply to all the school in a "window" around her expected outcome. Using Hoeffding's inequality, the real rank of a student is at most $x\sqrt{n}$ away from her expected rank, with probability at least $1-2e^{-2x^2}$. If schools have capacity c, then an upper quota of $\mathcal{O}(\sqrt{n}/c)$ applications is already large enough for students to ensure they get the same outcome they would obtain when applying to all the schools, with good probability. Such reasoning could be helpful to decision makers when setting an upper quota on the number of applications.

6.4 Existence of a Bayes-Nash equilibrium

6.4.1 Induced normal form

Following Milgrom and Weber [MW85], we define distributional strategies as the set $\tilde{\mathcal{S}}_\mu$ of probability distributions $\tilde{p} \in \Delta(T \times A)$ such that the marginal distribution on T is the distribution μ. In our setting, behavioral and distributional strategies are equivalent as there is a many-to-one mapping from a behavioral strategy p to the corresponding distributional strategy $\tilde{p}$.

- Given $\mu \in \Delta(T)$ and $p \in \mathcal{S}$, we define the distribution $\tilde{p}$ such that $\tilde{p}(B \times \{a\}) = \int_B p(t,a)\mathrm{d}\mu(t)$ for every Borel subset $B \subseteq T$ and for every action $a \in A$.
- Conversely, given $\tilde{p} \in \Delta(T \times A)$, first define the marginal distribution μ such that $\mu(B) = \tilde{p}(B \times A)$ for every Borel subset $B \subseteq T$. Then, for every action $a \in A$ the measure $B \mapsto \tilde{p}(B \times \{a\})$ is absolutely continuous with respect to μ, hence Radon-Nikodym theorem gives the existence of a measurable function $t \mapsto p(t,a)$ such that $\tilde{p}(B \times \{a\}) = \int_B p(t,a)\mathrm{d}\mu(t)$ for every $B \subseteq T$.

We now define the payoff function $\tilde{U}$ for distributional strategies. For every $(\tilde{p}_i)_{i\in[n]} \in \Delta(T \times A)^n$ we set

$$(\tilde{U}(\tilde{p}_i, \tilde{p}_{-i}))_{i\in[n]} = \frac{\sum_{a\in A^n} \int_{t\in T^n} \textsc{Utility}((t_i,a_i)_{i\in[n]}) \cdot \mathbb{1}[\text{all } t_i\text{'s are distinct}] \cdot \prod_{i\in[n]} \mathrm{d}\tilde{p}_i(t_i,a_i)}{\sum_{a\in A^n} \int_{t\in T^n} \mathbb{1}[\text{all } t_i\text{'s are distinct}] \cdot \prod_{i\in[n]} \mathrm{d}\tilde{p}_i(t_i,a_i)} \tag{6.1}$$

Notice that the transformation from behavioral to distributional strategies is payoff-preserving, that is

$$\forall (p_i)_{i\in[n]} \in \mathcal{S}^n, \quad (U_\mu(p_i, p_{-i}))_{i\in[n]} = (\tilde{U}(\tilde{p}_i, \tilde{p}_{-i}))_{i\in[n]}. \tag{6.2}$$

The *induced normal form* of the Bayesian game is the surrogate symmetric n players game $\mathcal{G}_\mu = \langle n, \tilde{\mathcal{S}}_\mu, \tilde{U}\rangle$, where the set of actions is the set of distributional strategies $\tilde{\mathcal{S}}_\mu$. Thus, a behavioral strategy profile $(p_i)_{i\in[n]}$ is a mixed Bayes-Nash equilibrium in the original Bayesian game if and only if the corresponding distributional strategy profile $(\tilde{p}_i)_{i\in[n]}$ is a pure Nash equilibrium in $\mathcal{G}_\mu$.

The notion of ε-equilibrium in the induced normal form game $\mathcal{G}_\mu$ exactly corresponds to *ex-ante* ε-equilibrium in the Bayesian game, that is a strategy profile where no student can deviate and win more than ε, in average before drawing her type.

6.4.2 Existence theorem

We are now ready to prove the existence of an equilibrium for the game $\mathcal{G}_\mu$, and thus for the Bayesian game.

Theorem 6.4. *The game $\mathcal{G}_\mu$ has a symmetric equilibrium.*

Proof. We apply Proposition 1, Proposition 3, and Theorem 1 from [MW85]. Because the set of actions A is finite, the game has equicontinuous payoffs (R1). Types of students are drawn without replacement from μ, which is equivalent with sampling types independently and condition on the fact that types are distinct. Thus the distribution over T^n (types drawn without replacement) is absolutely continuous with respect to the product distribution (types drawn independently), which proves that the game has absolutely continuous information (R2). Milgrom and Weber use Glicksberg's Theorem to prove that the best response correspondence has a fixpoint, which gives the existence of a Nash equilibrium. Proving the existence of a symmetric equilibrium only requires a small modification of the best response correspondence, see for example [Che+04]. □

6.5 Computing equilibria with finitely many types

In this section we provide algorithms to compute an equilibrium when μ has a finite support.

6.5.1 Symmetric agent form

When the distribution μ is discrete and has a finite support $\{t_1, \ldots, t_k\}$ of size $k \geq n$, we define the *symmetric agent-form* game $\mathcal{G}'_\mu = \langle k, A, V \rangle$, where each player corresponds to a type, and the payoff function $V : A^k \to \mathbb{R}^k_+$ is a vector-valued function, such that the i-th coordinate of $V(a)$ is equal to the expected payoff of a student of type t_i in the Bayesian game (the expectation is taken over the types of the $n-1$ other students) when a player of type t_j plays a_j. By construction, $p \in \mathcal{S}$ is a symmetric equilibrium of the Bayesian game if and only if $(p(t_i))_{i\in[k]} \in \Delta(A)^k$ is a mixed equilibrium of the agent-form game.

The notion of ε-equilibrium in the symmetric agent-form game $\mathcal{G}'_\mu$ exactly corresponds to *interim* ε-equilibrium in the Bayesian game, where no student can deviate and win more that ε after drawing her type (but before drawing the types of other players). Because an interim approximate equilibrium is also an ex-ante approximate equilibrium, any ε-equilibrium of the game $\mathcal{G}'_\mu$ induces an ε-equilibrium of the game $\mathcal{G}_\mu$.

Theorem 6.5. *For the game $\mathcal{G}'_\mu$, computing an exact equilibrium is in the class* FIXP*, and computing an approximate ε-equilibrium with $\varepsilon > 0$ is in the class* PPAD*.*

Proof. The game $\mathcal{G}'_\mu$ is a k-players game in normal form, where the payoff function V is given by a matrix of size $k \cdot |A|^k$. As such, the problems of computing exact and approximate equilibria are respectively in the classes FIXP and PPAD, see for example [Yan09]. □

6.5.2 One application per student and strong α-reducibility

The student-proposing deferred acceptance procedure being quite complex, we will make some assumption on the preferences of schools and students in order to simplify the matching procedure. More precisely, Algorithm 6.2 simplifies the function Utility when the matching market induced by $(t_i, a_i) \in (T \times A)^n$ is α-reducible.

Algorithm 6.2 Simplified procedure when the matching market is α-reducible.

Game parameters: n, m, A, $T = [0,1]^d$, $(s_j)_{j\in[m]}$, $(v_j)_{j\in[m]}$ and $(c_j)_{j\in[m]}$.
Initialization:
  Add a "sentinel" school 0 with capacity $c_0 = n$ and value $v_0 = 0$, ranked last in every list.
Function Utility$((t_i, a_i)_{i\in[n]} \in (T \times A)^n)$
  Each school j sorts its applicants ($i \in [n]$ such that $j \in a_i$) by decreasing score ($s_j(t_i)$).
  Initialize $r \leftarrow (c_j)_{0\leq j\leq m}$ the vector of remaining capacities.
  While some students are unassigned**, do**
    For each school j with a positive capacity ($r_j > 0$) and some unassigned applicant**, do**
      Let i be the top unassigned applicant at school j.
      If school j is student i's first choice among schools with a positive capacity**, then**
        Assign student i to school j, set $u_i \leftarrow v_j(t_i)$ and $r_j \leftarrow r_j - 1$.
  Return the vector of utilities $(u_i)_{i\in[n]}$.

Theorem 6.6. *If the matching market induced by $(t_i, a_i)_{i\in[n]} \in (T \times A)^n$ is α-reducible and schools give distinct scores to student, then Algorithm 6.2 is equivalent with Algorithm 6.1.*

Proof. Assuming that schools give distinct scores to students, the algorithm is able to sort applicants by decreasing scores. Assuming that the matching market is α-reducible, at least one student will be assigned at each iteration of the while loop and the algorithm will terminate. To show the equivalence with Algorithm 6.1, consider the first pair (i, j) assigned by Algorithm 6.2. By construction, i and j prefer each other to everyone else, thus must be matched in every stable matching, and in particular the one computed by the student proposing deferred acceptance algorithm. We continue with the same reasoning by induction. □

In the complete information setting, and when the matching market is α-reducible, Theorem 6.3 shows that every equilibrium implements the unique stable matching, and that one can build simple equilibrium by induction: there exist a fixed student-school pair (i, j) who prefer each other to everyone else, thus reporting j is a dominant strategy for player i, who can be removed from the market (together with her seat), and so on. In the incomplete information case, we build on this intuition, replacing a student-school pair by a type-school pair $(t, j) \in T \times [m]$.

Definition 6.7 (strong α-reducibility)**.** We say that the matching game is *strongly α-reducible*, where for every $S \subseteq T$ and $q \in [0, 1]^m$, there must be at least one pair $(t, j) \in S \times [m]$ such that $s_j(t') \leq s_j(t)$ for every $t' \in S$ and $q_{j'} \cdot v_{j'}(t) \leq q_j \cdot v_j(t)$ for every $j' \in [m]$.

As a special case, notice that if schools have identical preferences (all functions s_j are equal), or if students have identical preferences (all functions v_j are constant), or if preferences of students and schools are symmetric ($s_j = v_j$ for all j), then the game is strongly α-reducible.

Observe that strong α-reducibility does not ask for strict reducibility, because that would make the definition too restrictive. In a strongly α-reducible game, given each players' type and action, the resulting market is α-reducible if and only if each school give distinct scores to students.

In Algorithm 6.3, we compute an equilibrium of the symmetric agent-form game $\mathcal{G}'_\mu$. The algorithm can be implemented to run in time $\mathcal{O}(m \cdot k \cdot \ln k)$, which corresponds to the time needed for each school to sort types by decreasing score, and is much more efficient than a generic algorithm for the symmetric agent-form game with m actions and k players.

Algorithm 6.3 Compute an equilibrium with $\ell = 1$ and strong α-reducibility.

Game parameters: k, n, m, $A = [m]$, $T = [0, 1]^d$, $(s_j)_{j\in[m]}$, $(v_j)_{j\in[m]}$ and $(c_j)_{j\in[m]}$.
Function PROBA(x, c)
  Return $\sum_{i=0}^{c-1} \binom{x}{i}\binom{k-1-x}{n-1-i}/\binom{k-1}{n-1}$, that is, the probability that at most $c - 1$ of the other $n - 1$ students will draw one of the x types that have (already) been assigned to this school of capacity c.
Function EQUILIBRIUM(collection of types $(t_i)_{i\in[k]} \in T^k$)
  Initialize $x \leftarrow (0)_{j\in[m]}$ the number of types applying to each school.
  While some types are unassigned**, do**
    For each school $j \in [m]$**, do**
      Let i be an unassigned type which maximizes $s_j(t_i)$.
      If $v_{j'}(t_i) \cdot \text{PROBA}(x_{j'}, c_{j'}) \leq v_j(t_i) \cdot \text{PROBA}(x_j, c_j)$ for every other $j' \in [m]$**, then**
        Assign type i to school j, set $a_i \leftarrow j$ and $x_j \leftarrow x_j + 1$.
  Return the distributional strategy Uniform$\{(t_i, a_i)\}_{i\in[k]}$.

Theorem 6.8. *Let μ_k be the uniform distribution over $(t_i)_{i\in[k]}$. If each student is allowed $\ell = 1$ application, if every school gives distinct scores to types, and if the game is strongly α-reducible, then Algorithm 6.3 returns a symmetric equilibrium of the game $\mathcal{G}_{\mu_k}$ in time $\mathcal{O}(mk \ln k)$.*

Proof. First, we show that if the game is strongly α-reducible then Algorithm 6.3 terminates: at each iteration of the while loop there is at least one fixed pair with $q = (\textsc{Proba}(x_j, c_j))_{j\in[m]}$. Then we compute a symmetric equilibrium via the elimination of dominant strategies.

At each iteration, we define $q_j = \textsc{Proba}(x_j, c_j)$. Consider type-school pairs (t, j) where school j ranks t first among types that have not been assigned yet (there are no ties by assumption). Then, a student of type t will be accepted to school j with probability $= q_j$ (because she is ranked first in school j), and will be accepted to another school j' with probability $\le q_{j'}$ (because she might not be ranked first at school j'). If (t, j) is a fixed pair, then $q_{j'} \cdot v_{j'} \le q_j \cdot v_j$ for every j', and it is a dominant strategy for a student of type t to apply to school j.

A crucial detail is that student's types are drawn without replacement. This removes any feedback the strategy of a type may have on itself because of multiple students having the same type. This also explains why Proba uses an hypergeometric distribution rather than a simpler binomial distribution. □

6.5.3 Schools have identical preferences

Gusflied and Irving [GI89] observed that the matching is unique when all schools have identical preferences. In such a case, the matching procedure of Algorithm 6.2 further simplifies into the serial dictatorship mechanism: the best student chooses her favorite school, then the second best student chooses among remaining schools, and so on.

Algorithm 6.4 Computing an equilibrium when schools have identical preferences $s_j = s$.

Game parameters: n, m, A, $T = [0,1]^d$, s, $(c_j)_{j\in[m]}$ and $(v_j)_{j\in[m]}$.
Initialization:
  Add a "sentinel" school 0 with capacity $c_0 = n$ and value $v_0 = 0$, ranked last in every list.
  Define the set of "remaining capacity" vectors $R = \prod_{0\le j\le m}\{0, 1, \dots, c_j\}$.
Function School(preference list $a \in A$, remaining capacities $r \in R$)
  Return the first school j in the preference list a whose remaining capacity is $r_j > 0$.
Function Equilibrium(collection of types $(t_i)_{i\in[k]} \in T^k$)
  Sort types $(t_i)_{i\in[k]}$ by non-increasing order of score $s(t_i)$.
  Initialize the distribution $q \in \Delta(R)$ such that $q((c_j)_{0\le j\le m}) = 1$.
  For i from 1 to k, **do**
    Let $a_i = \arg\max_{a\in A} \sum_{r\in R} q(r) \cdot v_{\textsc{school}(a,r)}(t_i)$.
    For each "remaining capacity" vector $r \in R$ in lexicographical order, **do**
      Let $p \leftarrow (n - 1 - \sum_{0\le j\le m}(c_j - r_j))/(n - i + 1)$ be the probability
        that a student has type t_i.
      Let $r' \leftarrow r - \delta_j$ be the capacity vector once a student
        is assigned to school $j = \textsc{School}(a_i, r)$.
      Set $q(r) \leftarrow (1-p) \cdot q(r)$ and $q(r') \leftarrow q(r') + p \cdot q(r)$.
  Return the distributional strategy Uniform$\{(t_i, a_i)\}_{i\in[k]}$.

In Algorithm 6.4, we compute an equilibrium of the symmetric agent-form game $\mathcal{G}'_\mu$. The algorithm can be implemented to run in time $\mathcal{O}(k \cdot n \cdot \prod_{j\in[m]}(1 + c_j)\})$, which is linear in k, and thus much more efficient than a generic algorithm for the symmetric agent-form game with $\binom{m}{\ell}$ action for each of the k players.

Theorem 6.9. *Let μ_k be the uniform distribution over $(t_i)_{i\in[k]}$. If all school have the same scoring function s, and if all $s(t_i)$'s are distincts, then Algorithm 6.4 returns a symmetric equilibrium of the game $\mathcal{G}_{\mu_k}$.*

Proof. We compute a symmetric equilibrium by eliminating dominant strategies. After sorting types by decreasing scores, notice that the expected payoff of a student having type t_i does not depend on the strategy of students having types $t_{i+1}, \dots, t_k$.

At each iteration i, q corresponds to the distribution over remaining seats when the serial dictatorship mechanism consider a type t_i student (and assigns her to the first available school in her preference list). Notice that because we allow students to apply to more than 1 school, correlations may exist between the number of remaining seats in different schools, which is the reason why we store the whole distribution and not only its marginals. Then, we compute the expected payoff of each action, chose the best one, and update distribution q accordingly.

As in Theorem 6.8, types of students are drawn without replacement. This is exactly the distribution we consider when updating q: conditioning on the fact that remaining seats are given by r, exactly $x = \sum_{0\leq j\leq m}(c_j - r_j)$ students have drawn types in $\{t_1, \dots, t_{i-1}\}$, hence $n-1-x$ other students have types in $\{t_i, \dots, t_n\}$, and one of them will draw type t_i with probability $(n-1-x)/(n-i+1)$. □

6.6 Computing equilibria with an atomless type distribution

This section is a toolbox to prove that an algorithm approximates a Bayes Nash equilibrium. In Figures 6.9 and 6.10, we illustrate how one can combine algorithms from Section 6.5 with a convergence theorem to compute equilibria of games with a continuous distribution over types.

Weak convergence of type distribution. Computing an equilibrium is more tractable when the set of strategies has finite dimension. For that matter, when μ is continuous, we will discretize the set of types. Denote $\Delta^d(T) \subseteq \Delta(T)$ the set of discrete distributions having a finite support. Theorem 6.10 states that $\Delta^d(T)$ is dense in $\Delta(T)$ for the weak convergence of measures. More precisely, one can approximate a distribution by drawing a finite number of independent samples from it.

Theorem 6.10. *Let $\mu \in \Delta(T)$ be a distribution and let $(t_i)_{i\geq 1}$ be a sequence of independent random variables with distribution μ. For all $k \geq 1$, define the (random) distribution μ_k such that $\mu_k(B) = |\{t_i\}_{i\in[k]} \cap B|/k$ for all Borel set $B \subseteq T$. Then almost surely (over the randomness of the t_i's), the sequence $(\mu_k)_{k\geq 1}$ weakly converges towards μ.*

Proof. See [Var58]. □

Weak convergence of distributional strategies. Let us explain why the formalism of distributional strategies is required. For approximation purposes, we are interested in the case where each μ_k is discrete and has finite support, converging weakly towards a continuous distribution μ. For any behavioral strategy $p \in \mathcal{S}$, one could set each p_k to be equal to p almost everywhere (outside of the support of μ_k), while ensuring that each p_k is a symmetric Nash equilibrium under μ_k. Hence, assuming the weak convergence of behavioral strategies is not enough to prove that the limit is a Nash equilibrium.

Continuity of the payoff function. Unfortunately, assuming the weak convergence of a sequence of equilibria in distributional strategies, is still not enough to show that the limit is an equilibrium. In particular, we cannot directly apply Theorem 2 from Milgrom and Weber [MW85], because the payoff function $\textsc{Utility} : (T \times A)^n \to \mathbb{R}_+^n$ might be discontinuous in the players types. However, we will be able to show the weaker property that $\tilde{U} : \Delta(T \times A)^n \to \mathbb{R}_+$ is (sequentially) continuous at the limit (when distributional strategies are endowed with the topology of the weak convergence of measures). This requires additional continuity assumptions on μ, s_j's and v_j's.

Theorem 6.11. *If the following conditions hold, then the utility function $\tilde{U}$ is weakly continuous at every strategy profile in $(\tilde{\mathcal{S}}_\mu)^n$.*

- *the distribution μ is atomless (that is, $\mu(\{t\}) = 0$ for every $t \in T$),*
- *value and scoring functions are continuous μ-almost everywhere (that is, $\mu(D) = 0$ where D is the set of discontinuities of a scoring function s_j or of a value function v_j),*
- *level sets of scoring functions are μ-negligible (that is, $\mu(s_j^{-1}(\{y\})) = 0$ for every $j \in [m]$ and $y \in [0, 1]$).*

Proof. Let $(\tilde{p}_i)_{i\in[n]} \in (\tilde{\mathcal{S}}_\mu)^n$ be a strategy profile. In Equation (6.1) which defines $\tilde{U}$, we are going to show that integrands are continuous almost everywhere with respect to $(\tilde{p}_i)_{i\in[n]}$. Then, we conclude the proof using the Portmanteau Theorem (see for example Theorem 3.10.1 from [Dur19]), showing that $\tilde{U}$ is sequentially[4] continuous in $(\tilde{p}_i)_{i\in[n]}$.

First, assuming that μ is atomless is sufficient to prove that $\mathbb{1}[\text{all } t_i\text{'s are distinct}]$ is continuous almost everywhere. Moreover, if $\textsc{Utility}$ is not continuous in $(t_i, a_i)_{i\in[n]}$, then it is because a value function or a scoring function is discontinuous in some t_i, or because a school gives the same score to two types. Each condition occurs with probability 0, thus $\textsc{Utility}$ is continuous almost everywhere. □

Putting everything together. To approximate a Nash equilibrium of the game $\mathcal{G}_\mu$, first use Theorem 6.11 to show that $\tilde{U}$ is weakly continuous at every strategy profile in $(\tilde{\mathcal{S}}_\mu)^n$. Then use Theorem 6.10 to build a discrete approximation μ_k of the type distribution μ. Then, compute a symmetric Nash equilibrium $\tilde{p}_k$ of the game $\mathcal{G}_{\mu_k}$. Using Prokhorov's theorem, the set of distributional strategies $\Delta(T \times A)$ is metrizable and (sequentially) compact, hence one can build a converging subsequence of distributional strategies, whose limit will be a symmetric Nash equilibrium of $\mathcal{G}_\mu$.

Theorem 6.12. *Consider a sequence of measures $\mu_{k\geq n} \in \Delta(T)$ and a sequence of behavioral strategies $p_{k\geq n} \in \mathcal{S}$, if*

- *for all $k \geq n$, the distributional strategy $\tilde{p}_k \in \tilde{\mathcal{S}}_{\mu_k}$ is a symmetric equilibrium of $\mathcal{G}_{\mu_k}$,*
- *the sequence of distributional strategies weakly converges towards a strategy $\tilde{p} \in \tilde{\mathcal{S}}_\mu$ with a marginal type distribution $\mu \in \Delta(T)$,*
- *the payoff function $\tilde{U}$ is weakly continuous at every strategy profile in $(\tilde{\mathcal{S}}_\mu)^n$,*

then $\tilde{p}$ is a symmetric equilibrium for the game $\mathcal{G}_\mu = \langle n, \tilde{\mathcal{S}}_\mu, \tilde{U}\rangle$. Alternatively, if $\tilde{p}_k$'s are ε-equilibria with $\varepsilon > 0$ (ex-ante approximate equilibria of the Bayesian games), then $\tilde{p}$ is an ε-equilibrium of the game $\mathcal{G}_\mu$.

[4] Using Prokhorov's theorem, the space of probability measures $\Delta(T \times A)$ endowed with its weak topology is metrizable, thus the notions of sequential continuity and continuity are equivalent.

Proof. For the sake of contradiction, assume that $\tilde{p}$ is not a symmetric Nash equilibrium of $\mathcal{G}_\mu$. Then there exists a best response $\tilde{p}^* \in \tilde{\mathcal{S}}_\mu$ such that playing $\tilde{p}^*$ raises the payoff of the player by a positive constant $\varepsilon > 0$, that is $\tilde{U}(\tilde{p}^*, (\tilde{p})_{i\in[n-1]}) - \tilde{U}(\tilde{p}, (\tilde{p})_{i\in[n-1]}) = \varepsilon > 0$. The sequence of measures $(\mu_k)_{k\geq 0}$ weakly converges towards μ, hence there exists a sequence of distributional strategies $\tilde{p}_k^* \in \tilde{\mathcal{S}}_{\mu_k}$ weakly converging towards $\tilde{p}^*$. Using the continuity hypothesis on $\tilde{U}$, we show that $\tilde{U}(\tilde{p}_k^*, (\tilde{p}_k)_{i\in[n-1]}) - \tilde{U}(\tilde{p}_k, (\tilde{p}_k)_{i\in[n-1]})$ converges towards ε, and thus is positive for some $k \geq n$. Therefore, it contradicts the fact that each $\tilde{p}_k$ is an equilibrium of $\mathcal{G}_{\mu_k}$. The proof with approximate equilibria is identical, if we set ε in the proof to be equal to ε from the statement of the theorem. □

6.7 Simulations

For this chapter, implementations are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/Equilibrium

Convergence theorem with mixed equilibrium. Both Algorithms 6.3 and 6.4 compute a pure equilibrium $\tilde{p}_k$ of the game $\mathcal{G}_{\mu_k}$, in the sense that the behavioral strategy p_k is pure. Using Theorem 6.12 we show that $\tilde{p}_k$ weakly converge towards an equilibrium $\tilde{p}$ of $\mathcal{G}_\mu$. When p is a pure strategy, we can easily approximate p by the strategy p_k with a large k (see Figure 6.9). However, if p is a mixed strategy, we need an extra step to compute the limit: for every $t \in T$ we consider the $\sqrt{k}$ types from the support of μ_k that are closest to t, and let $p_k(t)$ be the average strategy over those points (see Figure 6.10).

Strongly α-reducible preferences. When the game is strongly α-reducible, we can compute equilibrium with $\ell = 1$ application per student. Special cases include when students have identical preferences (Figure 6.6) and when schools have identical preferences (Figure 6.7). In each case we implement Algorithm 6.3 in Python (see `identical-students.py` and `identical-schools.py` respectively), to generate Figure 6.9.

Schools have identical preferences. When schools have identical preferences, we can compute equilibrium for any $\ell \geq 1$. For simplicity our implementation also assumes that students have identical preferences. Because the complexity of Algorithm 6.4 is exponential in the number of students and the number of schools, an efficient implementation is preferable, which is the reason why we chose to have a Python script (`identical-all.py`) interacting with a C++ solver (`exact.cpp`).

The expensive part of Algorithm 6.4 is to evaluate the payoff of each action. To speed-up the computation, an improved solver (`approximate.cpp`) outputs an approximate equilibrium of $\mathcal{G}_{\mu_k}$, which will converge towards an approximate equilibrium of $\mathcal{G}_\mu$. The main idea is to replace the dynamic programming approach (where we compute q) by Monte Carlo simulations. We randomly partition $\{t_1, \ldots, t_k\}$ into $r = k/(n-1)$ sets of $n-1$ types, each corresponding to a "run". When considering the type t_i with $1 \leq i \leq k$, we approximate q by the empirical distribution of remaining capacities over the r runs. Figure 6.10 was obtained by setting $r = 2\,000\,000$ and $n = 50$, which runs in roughly 1 minute.

6.8 Conclusion and open questions

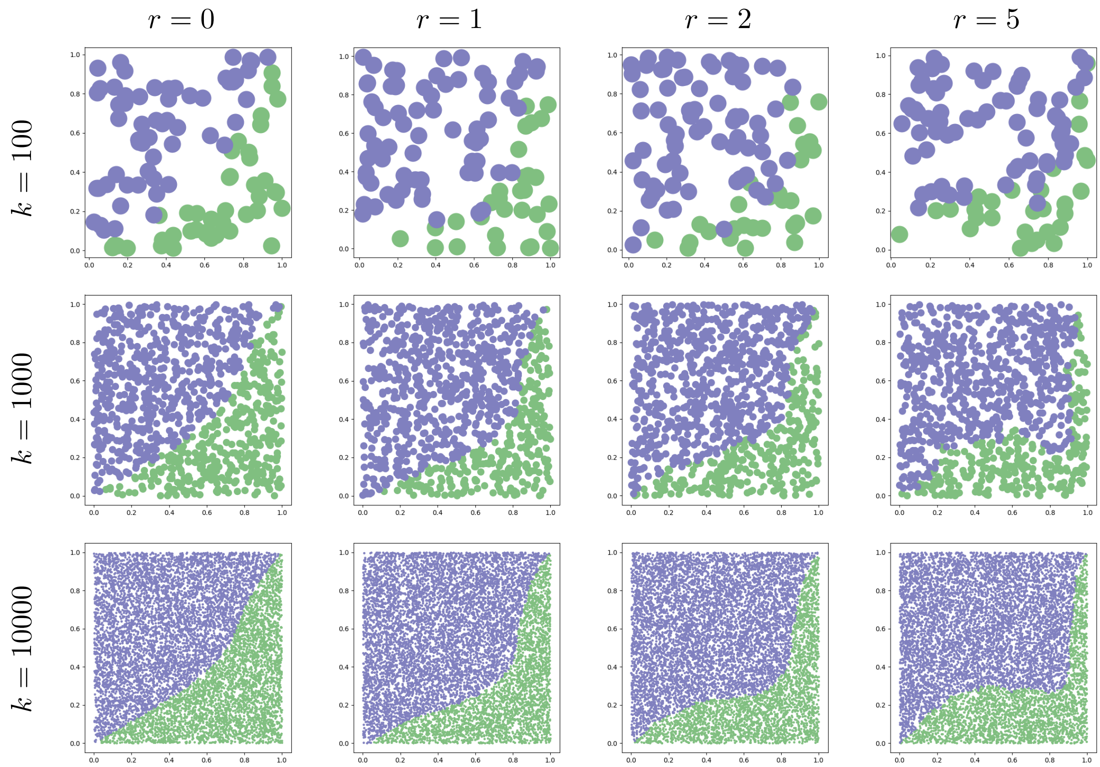
In this chapter, we generalized the game defined by Haeringer and Klijn [HK09], in a setting where students have incomplete information. We discussed the existence and the computability of equilibria in several setting. The following questions are left open for future work:

- **Equilibria with 1 application per student.** In the complete information case, Haeringer and Klijn show that equilibria with 1 application per student correspond to stable matchings. As illustrated in Figure 6.4, the incomplete information game can have an infinite number of equilibrium. But does the set of equilibria has a lattice structure?

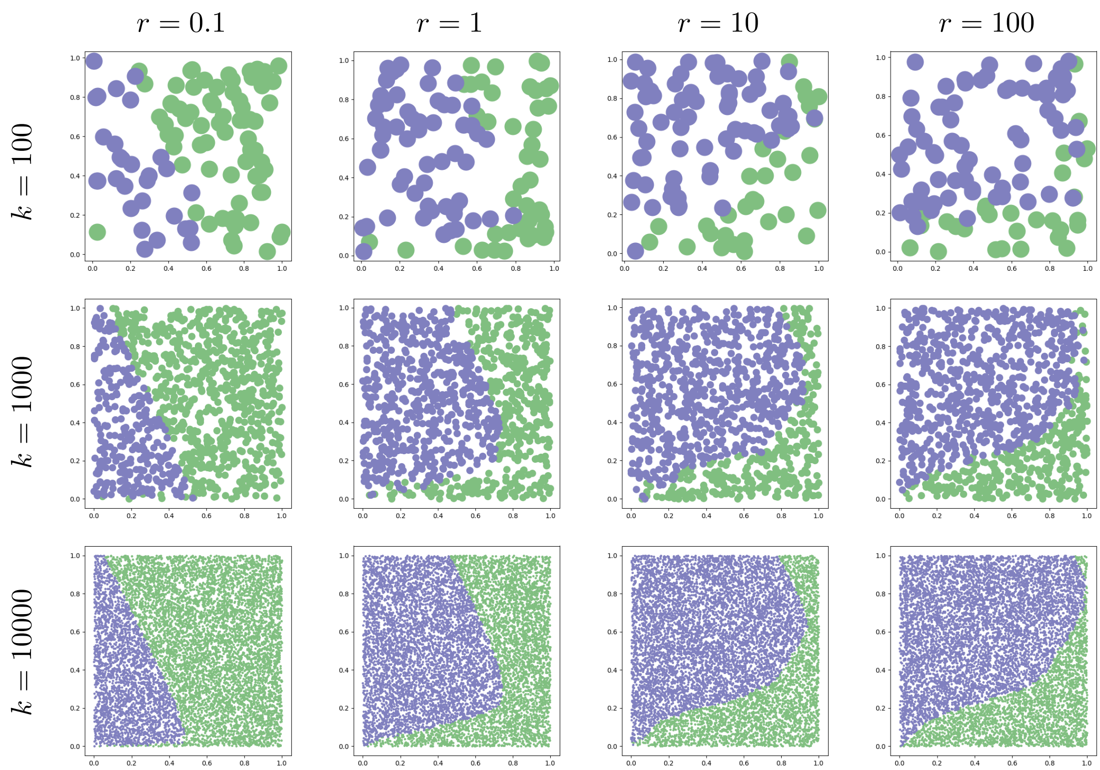
- **Unique equilibrium.** In Section 6.5, we compute equilibria with a finite number of types by eliminating dominant strategies. If each eliminated strategy strictly dominates other strategies, the equilibrium is unique. Using a convergence theorem, does uniqueness extends to the case where types are continuous?

- **Differential equations.** When combined with the convergence theorem, algorithms from Section 6.5 can be seen as first order Euler methods, which eventually solve differential equations. Such equations might lead to more efficient algorithms, and a to a proof that the equilibrium is unique.

- **Complexity.** In Section 6.5.1, we argue that a behavioral equilibrium with k types corresponds to an equilibrium in a k player game in normal form, and thus the problem of computing an exact/approximate equilibrium belong to the classes FIXP/PPAD. Are those results tight, in the sense of FIXP/PPAD-hardness?

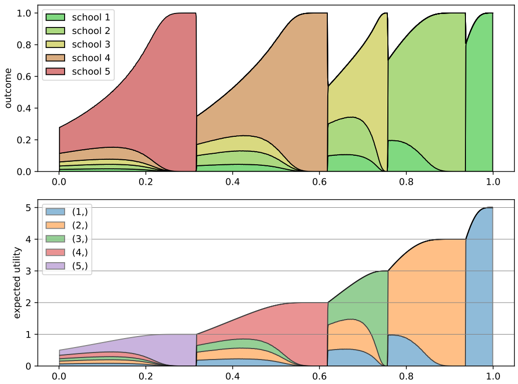


(a) Pure equilibria for random discretizations of the game from Figure 6.5.

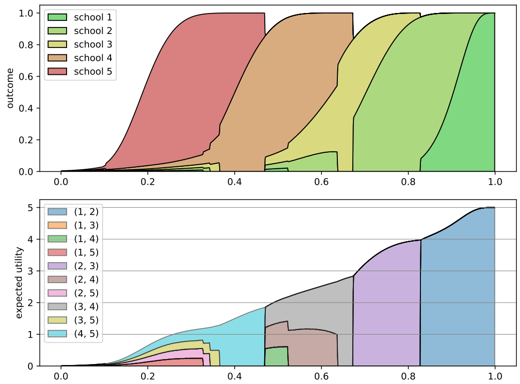


(b) Pure equilibria for random discretizations of the game from Figure 6.6.

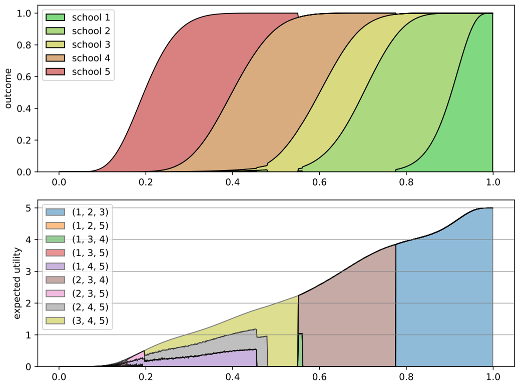
Figure 6.9. Equilibria with $\ell = 1$ for $\mathcal{G}_{\mu_k} = \langle n, \tilde{\mathcal{S}}_{\mu_k}, \tilde{U} \rangle$ computed using Algorithm 6.3, where μ_k is a distribution with a finite support of size k approximating μ. When $k \to +\infty$, the distributional strategies weakly converge towards the equilibria of $\mathcal{G}_\mu$ given in Figures 6.5 and 6.6.



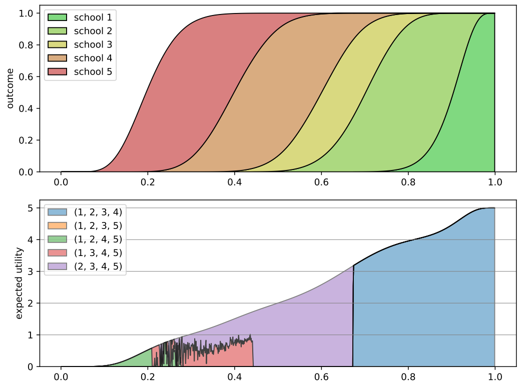
(a) Equilibrium with $\ell = 1$.



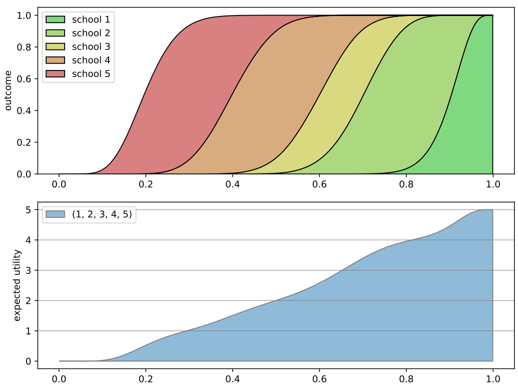
(b) Equilibrium with $\ell = 2$.



(c) Equilibrium with $\ell = 3$.



(d) Equilibrium with $\ell = 4$.



(e) Equilibrium with $\ell = 5$.

Figure 6.10. Equilibria of the game defined in Figure 6.8, computed using Algorithm 6.4. The top panel of each sub-figure plots the probability of each outcome, conditioned on the type of the student. In a mixed strategy, every action gives the same payoff. The bottom panel of each sub-figure decompose this payoff into the different strategy from the mixed equilibrium, conditioned on the type of the student. Exactly as we argued in Figure 6.8, the equilibrium with 3 applications per student yields almost the same outcome (in terms of expected payoff) as the one with 5 applications per student.

Part III

Who gets what?

7 | Output Distribution of Deferred Acceptance

This chapter is based on the following paper:

[Mau20] Simon Mauras. "Two-Sided Random Matching Markets: Ex-Ante Equivalence of the Deferred Acceptance Procedures". In: *Proceedings of the 21st ACM Conference on Economics and Computation*. 2020, pp. 585–597

7.1 Introduction

Gale and Shapley's deferred acceptance procedure comes with two variants: the men-proposing-deferred-acceptance (abbr. MPDA, see Algorithm 2.1), and the women-proposing-deferred-acceptance (abbr. WPDA). Consider a random two-sided matching market, where a (given) procedure computes a stable matching. Every agent of the market is interested by the distribution of outcomes. But computing which outcome an agent can expect is a difficult question, that has only been answered in special cases (for example, see Theorem 4.19 in Chapter 4 when both sides of the market have vertical preferences). Trying to answer the question "who gets what?", we discovered numerically an intriguing mathematical property which holds in the vanilla model where agents have incomplete uniform preferences: procedures MPDA and WPDA are *ex-ante equivalent*, in the sense that they induce the exact same distribution over matchings (see Section 7.2 and Theorem 7.2).

We prove this property using the lattice structure of stable matchings: an application of the inclusion-exclusion principle gives the probability that a matching is men/women-optimal (see Section 7.3 and Lemma 7.3). First, we show that the ex-ante equivalence property remains valid in the larger class of symmetric anti-popularity preference distributions (see Section 7.4 and Theorem 7.5). Then, we study the robustness of our result under other input distributions: symmetric popularity preferences (see Section 7.5); general utility preferences (see Section 7.7); and general anti-popularity preferences (see Section 7.8). In the latter case, Theorem 7.11 gives a closed formula for the probability of two person being matched, in balanced markets where agents have complete anti-popularity preferences, answering our initial question.

Related works. Although in the setting of one-sided matching markets, papers have shown the equivalence of random mechanisms under deterministic inputs, which can be interpreted as the equivalence of deterministic mechanisms under random inputs. Both Knuth [Knu96], and Abdulkadiroglu and Sönmez [AS98] show that in housing markets, computing the core allocation with random endowments is equivalent to the random serial dictatorship mechanism.

In two-sided matching markets, the ex-ante equivalence property can be compared to the stronger *ex-post equivalence*, where both MPDA and WPDA output the same stable matching. Most of the literature focus on asymptotic ex-post equivalence, under a large market assumption. In particular, Immorlica and Mahdian [IM15], and Ashlagi, Kanoria and Leshno [AKL17] show that the fractions of agents who do not receive the same allocation from MPDA and WPDA vanishes with a large number of agents.

Other articles from the literature of matching under random preferences have studied who gets what. Lee [Lee16] considers a model where agents have random vertical utility preferences, and shows that in every stable matching, agents asymptotically receive utilities equal to their public values. More recently, Ashlagi, Braverman, Saberi, Thomas and Zhao [Ash+21] show that the output distribution of deferred acceptance is asymptotically uniform when agent have aligned popularity preferences, with bounded popularities.

Takeaway message. Recall that MPDA outputs the men-optimal-stable-matching, and that WPDA outputs the women-optimal-stable-matching. Hence, the two procedures output the same matching when men and women like the same things, or more precisely when people like persons who like them in return. The same remark with random preferences give an intuitive explanation of our result: if men and women like the same things in average (symmetric preferences), then MPDA and WPDA will output the same matchings in average (ex-ante equivalence).

7.2 Motivating special case

The starting point of this work was to understand the output distributions of MPDA and WPDA, in a very simple matching market with M men and W women having heterogeneous preferences (agents have idiosyncratic preferences).

Definition 7.1 (Incomplete uniform preferences)**.** Consider any fixed bipartite graph $G = (\mathcal{M} \cup \mathcal{W}, E)$ with $\mathcal{M} = \{m_1, \dots, m_M\}$ the set of men, $\mathcal{W} = \{w_1, \dots, w_W\}$ the set of women, and $E \subseteq \mathcal{M} \times \mathcal{W}$ the set of edges. Agents have uniform preferences (see Definition 3.1) and rank their neighbours (non-edges are not acceptable) uniformly and independently at random.

Figure 7.1 illustrates Definition 7.1, on a bipartite graph with 5 men and 4 women. The output distributions of procedures MPDA and WPDA can be computed with a computer, and they happen to be identical. They are given in Figure 7.2.

For every bipartite graph with $M, W \leq 4$, we used computer simulations to compute the output distribution of MPDA and WPDA. Surprisingly the two output distributions were always identical, which led us to conjecture Theorem 7.2.

Theorem 7.2. *In a random matching market where the preference profile is drawn from an incomplete uniform preference distribution, the output distributions of MPDA and WPDA are identical.*

Proof sketch. In Section 7.4 we prove Theorem 7.5, which generalizes Theorem 7.2. Nonetheless, let us give the ideas of the proof on the example of Figures 7.1 and 7.2.

To compute the probability that WPDA outputs μ_1, we first compute the probability that μ_1 is stable. Then we subtract the probability that μ_1 is stable but not women-optimal, because another matching is stable and improves the outcome of women.

Each person is either matched in all stable matchings, or single in all stable matchings [MW70]; hence when μ_1 is stable, the only other matchings which can be stable are μ_2, μ_3, μ_4 and μ_5. Using the fact that stable matchings have a lattice structure [Knu76; Knu97], one can prove that when μ_1 is stable, the set of matchings that are stable and preferred to μ_1 by all women is either $\emptyset$, $\{\mu_2\}$,

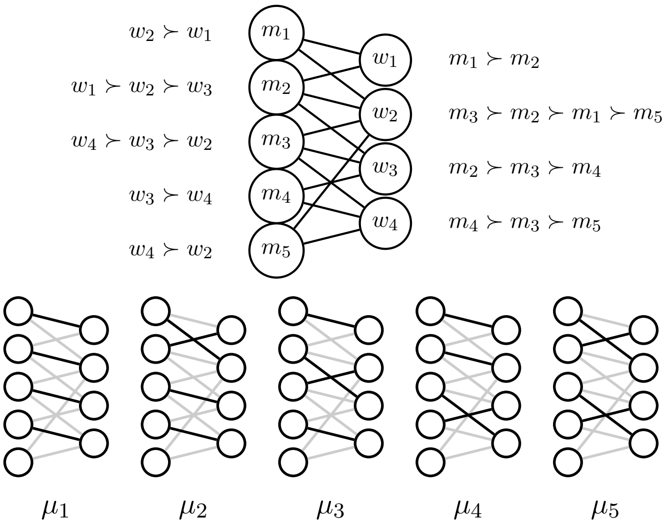


Figure 7.1. Example of incomplete uniform preference distribution. The probability of sampling this particular preference profile is $1/(2!^4 \cdot 3!^4 \cdot 4!) = 1/497664$. There are five stable matchings μ_1, μ_2, μ_3, μ_4 and μ_5. MPDA outputs μ_5 and WPDA outputs μ_3.

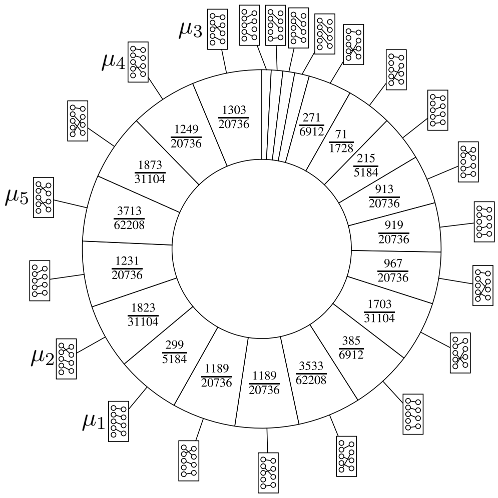


Figure 7.2. Output distribution, common to the procedures MPDA and WPDA, using the input model of Figure 7.1. The support of the distribution is the set of all maximal matchings.

$\{\mu_3\}$, $\{\mu_4\}$, $\{\mu_2,\mu_5\}$, $\{\mu_3,\mu_5\}$, $\{\mu_4,\mu_5\}$ or $\{\mu_2,\mu_4,\mu_5\}$. WPDA outputs μ_1 in the first case, which occurs with a probability that can be computed with an inclusion-exclusion principle.

$$\begin{aligned}\mathbb{P}[\text{WPDA outputs } \mu_1] = \mathbb{P}[\mu_1 \text{ stable}] &- \mathbb{P}[\mu_1 \text{ and } \mu_2 \text{ stable, women prefer } \mu_2]\\ &- \mathbb{P}[\mu_1 \text{ and } \mu_3 \text{ stable, women prefer } \mu_3]\\ &- \mathbb{P}[\mu_1 \text{ and } \mu_4 \text{ stable, women prefer } \mu_4]\\ &+ \mathbb{P}[\mu_1, \mu_2 \text{ and } \mu_4 \text{ stable, women prefer } \mu_2 \text{ and } \mu_4]\end{aligned}$$

It turns out that each probability on the right-hand side of the equality is equal to its counterpart where we swap the roles of men and women. This can be shown using integral formulae defined by Pittel [Pit92].

$$\begin{aligned}\mathbb{P}[\mu_1 \text{ and } \mu_2 \text{ stable, women prefer } \mu_2] &= \int_0^1 \cdots \int_0^1 \mathrm{d}\mathbf{x}_1 \cdot \mathrm{d}\mathbf{x}_2 \cdot \mathrm{d}\mathbf{x}_3 \cdot \mathrm{d}\mathbf{x}_4 \cdot \mathrm{d}\mathbf{y}_1 \cdot \mathrm{d}\mathbf{y}_2 \cdot \mathrm{d}\mathbf{y}_3 \cdot \mathrm{d}\mathbf{y}_4\\ &\quad\cdot (1-\mathbf{x}_2\mathbf{y}_3)\cdot(1-\mathbf{x}_3\mathbf{y}_2)\cdot(1-\mathbf{x}_3\mathbf{y}_4)\cdot(1-\mathbf{x}_4\mathbf{y}_3)\\ &\quad\cdot(1-\mathbf{y}_2)\cdot(1-\mathbf{y}_4)\cdot\mathbf{x}_1\cdot\mathbf{x}_2\cdot\mathbf{y}_1\cdot\mathbf{y}_2\\ &= \mathbb{P}[\mu_1 \text{ and } \mu_2 \text{ stable, men prefer } \mu_2]\end{aligned}$$

Swapping the role of men and women in the left-hand side of the equality, gives the probability that MPDA outputs μ_1, hence MPDA and WPDA are equally likely to output μ_1. As a sanity check, we can compute the output probability of μ_1, and check that this theoretical value matches the experimental one from Figure 7.2.

$$\mathbb{P}[\text{WPDA outputs } \mu_1] = \frac{2795}{41472} - \frac{437}{124416} - \frac{7}{2592} - \frac{19}{5184} + \frac{5}{31104} = \frac{299}{5184}$$

For every fixed matching, the same type of arguments apply, which concludes the proof. □

7.3 Main Lemma: Inclusion-exclusion principle

In this section, we compute the probability that a matching μ is the men/women-optimal stable matching. To do so, we use an inclusion-exclusion principle on the set of rotations which could be exposed and women/men-improving in μ.

Lemma 7.3. *Consider a matching market with random preferences. The probability that a matching $\mu : \mathcal{M}\cup\mathcal{W} \to \mathcal{M}\cup\mathcal{W}$ is stable and men/women-optimal is*

$$\mathbb{P}[\mu \textit{ is stable and women-optimal}] = \sum_{\substack{\sigma \textit{ permutation}\\ \sigma_{|\mathcal{M}}=\mu_{|\mathcal{M}}}} (-1)^{C(\sigma)} \cdot \mathbb{P}[\sigma \textit{ is stable}]$$

$$\mathbb{P}[\mu \textit{ is stable and men-optimal}] = \sum_{\substack{\sigma \textit{ permutation}\\ \sigma_{|\mathcal{W}}=\mu_{|\mathcal{W}}}} (-1)^{C(\sigma)} \cdot \mathbb{P}[\sigma \textit{ is stable}]$$

where $C(\sigma)$ is the number of cycle of length > 2 in σ.

Proof. The men and women cases being symmetric, we prove the formula giving the probability that a matching is stable and women-optimal. Let $\mathcal{R}$ be the set of rotations r such that $r(m) = \mu(m)$ for all man $m \in r$. Matching μ is outputted by WPDA when it is stable and women-optimal: no rotation $r \in \mathcal{R}$ is exposed and women-improving in μ.

$$\mathbb{P}[\mu \text{ is stable and women-optimal}] = \mathbb{P}[\mu \text{ is stable}] - \mathbb{P}[\mu \text{ is stable and some } r \in \mathcal{R} \text{ is exposed}]$$

Using an inclusion-exclusion principle to compute the probability of a disjunction, we obtain:

$$\mathbb{P}[\mu \text{ is stable and women-optimal}] = \sum_{R \subseteq \mathcal{R}} (-1)^{|R|} \cdot \mathbb{P}[\mu \text{ is stable and every } r \in R \text{ is exposed}]$$

Recall that two different rotations can be exposed at the same time only if they are disjoint. Thus, we can consider only sets $R \subseteq \mathcal{R}$ of disjoint rotations. Moreover, μ is stable and every rotation from R is exposed if an only if the associated permutation σ_R is stable.

$$\sigma_R : \begin{cases} m \mapsto \mu(m) & \text{if } m \in \mathcal{M} \\ w \mapsto \mu(w) & \text{if } w \in \mathcal{W} \text{ and } w \notin r \text{ for all } r \in R \\ w \mapsto r(w) & \text{if } w \in \mathcal{W} \text{ and } w \in r \text{ for some } r \in R \end{cases}$$

If $C(\sigma)$ is the number of cycles of length > 2 in σ, we have $C(\sigma_R) = |R|$, concluding the proof. □

7.4 Input model: Symmetric anti-popularity preferences

After observing that the output distributions of MPDA and WPDA are identical when the preference profile is generated from a bipartite graph (see Section 7.1), we used computer simulations on more general classes of input distributions. We observed that MPDA and WPDA are ex-ante equivalent with the input model illustrated in Figure 7.3 and defined in Definition 7.4.

$P =$

	w_1	w_2	w_3
m_1	2	4	0
m_2	8	1	4

The preference list of m_2 is $w_2 \succ w_1 \succ w_3$ with probability:

$$\frac{1/4}{1/1+1/4+1/8} \cdot \frac{1/8}{1/1+1/8} \cdot \frac{1/1}{1/1} \approx 0.02$$

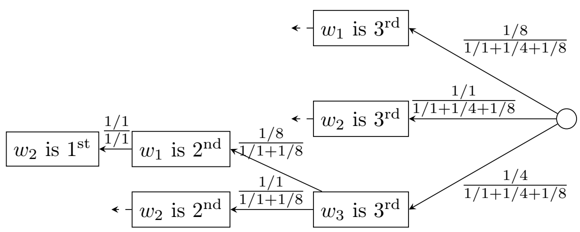


Figure 7.3. Symmetric anti-popularity preference distribution

Definition 7.4 (Symmetric anti-popularity preferences)**.** When men and women have symmetric (Definition 3.11) antipopularity (see Definition 3.3) preferences, popularities are given by a function $P : \mathcal{M} \times \mathcal{W} \to \mathbb{R}_+$, where $P(m, w)$ is the "popularity" that m and w attribute to each other, and where pairs with popularity 0 are not acceptable.

We say that this preference distribution is symmetric because the "popularity" that m gives to w is the same as the "popularity" that w gives to m. The "popularity" parameter $P(m, w)$ relates to how likely are m and w to like each other. In particular, a man m will prefer woman w_1 to woman w_2 with probability $P(m, w_1)/(P(m, w_1) + P(m, w_2))$.

Theorem 7.5. *In a random matching market where the preference profile is drawn from a symmetric anti-popularity preference distribution, the output distributions of MPDA and WPDA are identical.*

Proof. A fixed matching μ is outputted by WPDA (resp. MPDA) if and only if it is stable and women-optimal (resp. men-optimal). In Lemma 7.3, we give a formula for the probability that μ is stable and men/women-optimal. Moreover, for every permutation σ we have:

- σ and σ^{-1} are equally likely to be stable (proved in the upcoming Lemma 7.6).
- σ and σ^{-1} have the same number of cycles of length > 2 (that is $C(\sigma) = C(\sigma^{-1})$).
- $\sigma_{|\mathcal{M}} = \mu_{|\mathcal{M}}$ if and only if $\sigma^{-1}_{|\mathcal{W}} = \mu_{|\mathcal{W}}$

The two sums from Lemma 7.3 are equal by re-indexing, thus $\mathbb{P}[\mu$ is stable and women-optimal$] = \mathbb{P}[\mu$ is stable and men-optimal$]$. The matching μ has the same probability of being the output of MPDA and WPDA, which concludes the proof. □

Lemma 7.6. *When men and women have symmetric anti-popularity preferences, a permutation $\sigma : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W}$ and its inverse σ^{-1} are equally likely to be stable.*

Proof. If some person does not find acceptable their successor or their predecessor, then neither σ nor σ^{-1} can be stable, thus both probabilities are equal to 0. In the following, we assume that all matches induced by σ and σ^{-1} are acceptable. To compute the probability that a permutation σ is stable, we analyze the following randomized algorithm:

1. For each person x, we (partially) draw their preference list using anti-popularities, starting from the least favorite partner, and stopping as soon as we see either $\sigma(x)$ or $\sigma^{-1}(x)$.
2. If for some acceptable man-woman pair (m, w), we have not seen m when drawing w's preferences, nor have we seen w when drawing m's preferences, then σ is not stable.
3. If for some person x such that $\sigma(x) \neq \sigma^{-1}(x)$, we stopped after drawing $\sigma(x)$, then x prefers their predecessor $\sigma^{-1}(x)$ to their successor $\sigma(x)$, and the permutation σ is not stable.

Recall that a pair (m, w) is blocking the permutation σ if both m and w prefer each other to their predecessors. If the procedure described above does not fail in steps (2) or (3), then σ is stable.

Analyzing this procedure requires careful handling of conditional probabilities. For each person x, we condition on the identity and ordering of partners to whom x prefers both $\sigma(x)$ and $\sigma^{-1}(x)$, but we do not condition on whether x prefers $\sigma(x)$ to $\sigma^{-1}(x)$. Using this conditioning, steps (1) and (2) are deterministic, and their outcome are identical when checking the stability of σ and σ^{-1}. Then, using the definition of anti-popularity preferences, one can compute the probability that each person prefers their successor to their predecessor. Let us start using the example from Figure 2.2. In permutation σ, each person of the cycle $m_2 \mapsto w_2 \mapsto m_3 \mapsto w_3 \mapsto m_2$ prefer their successor to their predecessor with probability

$$\underbrace{\frac{P(m_2, w_2)}{P(m_2, w_3) + P(m_2, w_2)}}_{\mathbb{P}[\sigma(m_2) \succ_{m_2} \sigma^{-1}(m_2)]} \cdot \underbrace{\frac{P(m_3, w_3)}{P(m_3, w_2) + P(m_3, w_3)}}_{\mathbb{P}[\sigma(m_3) \succ_{m_3} \sigma^{-1}(m_3)]} \cdot \underbrace{\frac{P(m_3, w_2)}{P(m_2, w_2) + P(m_3, w_2)}}_{\mathbb{P}[\sigma(w_2) \succ_{w_2} \sigma^{-1}(w_2)]} \cdot \underbrace{\frac{P(m_2, w_3)}{P(m_3, w_3) + P(m_2, w_3)}}_{\mathbb{P}[\sigma(w_3) \succ_{w_3} \sigma^{-1}(w_3)]}.$$

However, in the inverse permutation σ^{-1} we reverse every edge of the cycle. Observe that the probability that each person prefer their successor to their predecessor remains the same.

$$\underbrace{\frac{P(m_2, w_3)}{P(m_2, w_3) + P(m_2, w_2)}}_{\mathbb{P}[\sigma^{-1}(m_2) \succ_{m_2} \sigma(m_2)]} \cdot \underbrace{\frac{P(m_3, w_2)}{P(m_3, w_2) + P(m_3, w_3)}}_{\mathbb{P}[\sigma^{-1}(m_3) \succ_{m_3} \sigma(m_3)]} \cdot \underbrace{\frac{P(m_2, w_2)}{P(m_2, w_2) + P(m_3, w_2)}}_{\mathbb{P}[\sigma^{-1}(w_2) \succ_{w_2} \sigma(w_2)]} \cdot \underbrace{\frac{P(m_3, w_3)}{P(m_3, w_3) + P(m_2, w_3)}}_{\mathbb{P}[\sigma^{-1}(w_3) \succ_{w_3} \sigma(w_3)]}$$

This holds for every cycle in every permutation. Thus, our procedure has the same probability of failure when checking the stability of σ and σ^{-1}, which concludes the proof. □

7.5 Counter example: Symmetric popularity preferences

The input model described in the previous section is very similar (but not equivalent) to the model studied in [IM15; KP09]: agents build their preference lists by sampling without replacement from a distribution (in [IM15; KP09] agents first draw their favorite partner, in the previous section agents first draw their least preferred partner). A natural question is whether the ex-ante equivalence property holds with the more classical symmetric popularity distributions, illustrated in Figure 7.4. In Figure 7.5, we show that the answer to this question is no.

$P =$

	w_1	w_2	w_3
m_1	2	4	0
m_2	8	1	4

The preference list of m_2 is $w_2 \succ w_1 \succ w_3$ with probability:

$$\frac{1}{1+4+8} \cdot \frac{8}{4+8} \cdot \frac{4}{4} \approx 0.05$$

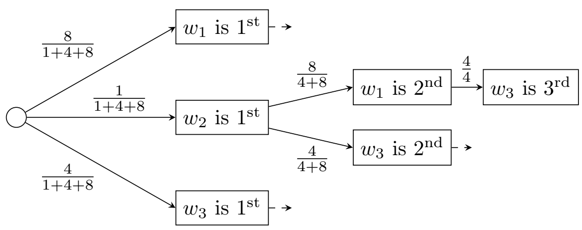


Figure 7.4. Symmetric popularity preference distribution

Definition 7.7 (Symmetric popularity preferences)**.** When men and women have symmetric (Definition 3.11) popularity (see Definition 3.2) preferences, popularities are given by a function $P : \mathcal{M} \times \mathcal{W} \to \mathbb{R}_+$, where $P(m, w)$ is the "popularity" that m and w attribute to each other, and where pairs with popularity 0 are not acceptable.

The formal definition of a *symmetric popularity preference distribution* is nearly identical to Definition 7.4, but preference lists are built from the start, drawing partners without replacement with probability proportional to their popularity. Both in the anti-popularity model (see Figure 7.3) and in the popularity model (see Figure 7.4), a man m will prefer woman w_1 to woman w_2 with probability $P(m, w_1)/(P(m, w_1) + P(m, w_2))$. The difference between the two distributions comes from rare events: in the anti-popularity setting a very popular person will sometimes be ranked unusually low, whereas in the popularity setting a very unpopular person will sometimes be ranked unusually high.

Popularity matrix P_ω, with $\omega > 1$.

$P_\omega =$

	w_1	w_2	w_3
m_1	ω	ω^2	0
m_2	ω^3	1	ω^2

Input model when $\omega = 2$:

- Anti-popularity → Figure 7.3
- Popularity → Figure 7.4

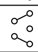
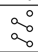
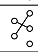
	Anti-popularity MPDA & WPDA	Popularity MPDA	Popularity WPDA
	$\sim 1/\omega^5$	$\sim 1/\omega^4$	$\sim 3/\omega^5$
	$\sim 1/\omega^2$	$\sim 1/\omega^2$	$\sim 1/\omega^2$
	$\sim 1/\omega$	$\sim 1/\omega$	$\sim 1/\omega$
	~ 1	~ 1	~ 1

Figure 7.5. Comparing the output distributions of MPDA and WPDA.

Figure 7.5 compares the output distributions of MPDA and WPDA, in both the popularity and anti-popularity setting, where popularities are parameterized by an arbitrary large constant $\omega > 1$. From Theorem 7.5, we know that the ex-ante equivalence property holds in the anti-popularity setting, that is the output of MPDA and WPDA are identical. In the popularity setting, we show that some matching can be arbitrarily more likely to be chosen by one of the two deferred-acceptance procedures. However, such matching is unlikely to be chosen by either procedures.

More precisely, in the popularity setting of Figure 7.5, MPDA matches m_2 and w_2 with probability $\sim 1/\omega^4$, whereas WPDA matches m_2 and w_2 with probability $\sim 3/\omega^5$. This is in part due to the fact that if $w_2 \succ_{m_2} w_1 \succ_{m_2} w_3$ (which occurs with probability $\sim 1/\omega^3$), MPDA matches m_2 and w_2 with probability $\sim 1/\omega$, whereas WPDA matches m_2 and w_2 with probability $\sim 1/\omega^2$.

7.6 Previous results: probability of stability

Random matching markets with N men and N women having complete uniformly random preference lists were studied in [Knu76; Knu97; Pit89; Pit92]. Knuth gave an integral formula for the probability p_N that a fixed matching is stable; with the objective of computing the asymptotic average number of stable matchings (in the uniform case, all $N!$ matchings have the same probability of being stable). In 1989, Pittel gave an alternate proof of this integral formula, and showed that $N! \cdot p_N \sim e^{-1} N \ln N$.

Let us retranscribe Pittel's proof of the integral formula. Let μ be any matching. Let X and Y be two random matrices, uniformly drawn from $[0,1]^{\mathcal{M}\times\mathcal{W}}$. Man m prefers woman w_1 to woman w_2 if $X_{m,w_1} < X_{m,w_2}$. Correspondingly, woman w prefers man m_1 to man m_2 if $Y_{m_1,w} < Y_{m_2,w}$.

Thus, a pair (m,w) is blocking matching μ if and only if $X_{m,w} < X_{m,\mu(m)}$ and $Y_{m,w} < Y_{\mu(w),m}$. We condition on the values of $\mathbf{x} = [X_{m,\mu(m)}]_{m\in\mathcal{M}}$ and $\mathbf{y} = [Y_{\mu(w),w}]_{w\in\mathcal{W}}$, and write the probability that a pair blocks μ:

$$\forall (m,w) \text{ such that } \mu(m) \neq w \text{ and } \mu(w) \neq w, \quad \mathbb{P}[(m,w) \text{ blocks } \mu \,|\, \mathbf{x},\mathbf{y}] = \mathbf{x}_m \cdot \mathbf{y}_w$$

Still conditioning on $\mathbf{x}$ and $\mathbf{y}$, blocking events are independent, hence the formula:

$$\mathbb{P}[\mu \text{ is stable}] = \underbrace{\int\cdots\int}_{2N} \mathrm{d}\mathbf{x}\cdot\mathrm{d}\mathbf{y}\cdot \prod_{\substack{m,w\\ \mu(m)\neq w\\ \mu(w)\neq m}} (1-\mathbf{x}_m\mathbf{y}_w)$$

In subsequent works [PI94; Pit19], Pittel extended the above formula to compute the probability that a fixed permutation is stable. We recall that a permutation σ is stable if the following is true:

- Every person x prefers their successor to their predecessors ($\sigma(x) \succeq_x \sigma^{-1}(x)$)
- For each pair $(m,w) \in \mathcal{M}\times\mathcal{W}$, we have ($\sigma^{-1}(m) \succeq_m w$) or ($\sigma^{-1}(w) \succeq_w m$)

We condition on the values of $\mathbf{x} = [X_{m,\sigma^{-1}(m)}]_{m\in\mathcal{M}}$ and $\mathbf{y} = [Y_{\sigma^{-1}(w),w}]_{w\in\mathcal{W}}$.

- Each man m such that $\sigma(m) \neq \sigma^{-1}(m)$ prefers $\sigma(m)$ to $\sigma^{-1}(m)$ with probability $\mathbf{x}_m$.
- Each woman w such that $\sigma(w) \neq \sigma^{-1}(w)$ prefers $\sigma(w)$ to $\sigma^{-1}(w)$ with probability $\mathbf{y}_w$.
- Each pair (m,w) such that $\sigma(m) \neq w$ and $\sigma(w) \neq m$ is blocking with probability $\mathbf{x}_m\mathbf{y}_w$.

Hence the formula:

$$\mathbb{P}[\sigma \text{ is stable}] = \underbrace{\int \cdots \int}_{2N} \mathrm{d}\mathbf{x} \cdot \mathrm{d}\mathbf{y} \cdot \prod_{\substack{m,w \\ \sigma(m)=w \\ \sigma(w)\neq m}} \mathbf{x}_m \cdot \prod_{\substack{m,w \\ \sigma(m)\neq w \\ \sigma(w)=m}} \mathbf{y}_w \cdot \prod_{\substack{m,w \\ \sigma(m)\neq w \\ \sigma(w)\neq m}} (1-\mathbf{x}_m\mathbf{y}_w)$$

For the more general problem of stable roommates, Mertens [Mer15] combined this formula with an inclusion-exclusion principle to compute the probability that a random instance has a solution.

7.7 Complete utility preferences

In this section, we define utility preference distributions, and generalize Pittel's integral formula to this setting. In Theorem 7.9, we derive a new integral formula for the probability that a matching is stable and men/women optimal. For simplicity, in this section we consider balanced matching markets with complete preferences.

Definition 7.8 (Complete utility preferences)**.** When men and women have complete (all pairs are acceptable) utility (Definition 3.5) preferences, we define $U_{m,w}$ the utility that man m gets if he is matched with w, and $V_{m,w}$ the utility that woman w gets if she is matched with m, such that $U_{m,w}$ and $V_{m,w}$ are independent continuous random variables on $\mathbb{R}_+$, with survival functions $S_{m,w} : u \mapsto \mathbb{P}[U_{m,w} > u]$ and $T_{m,w} : v \mapsto \mathbb{P}[V_{m,w} > v]$.

Theorem 7.9. *Consider a balanced matching market with complete utility preferences, defined by the survival functions $(S_{m,w})$ and $(T_{m,w})$. Then, for every matching μ, we have*[a]

$$\mathbb{P}[\mu \textit{ is stable}] = \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} \mathrm{d}\mathbf{u} \cdot \mathrm{d}\mathbf{v} \cdot \Phi^\mu(\mathbf{u},\mathbf{v})$$

$$\mathbb{P}[\mu \textit{ is stable and men-optimal}] = \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} \mathrm{d}\mathbf{u} \cdot \mathrm{d}\mathbf{v} \cdot \Phi^\mu(\mathbf{u},\mathbf{v}) \cdot \mathcal{T}^\mu(\mathbf{u},\mathbf{v})$$

$$\mathbb{P}[\mu \textit{ is stable and women-optimal}] = \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} \mathrm{d}\mathbf{u} \cdot \mathrm{d}\mathbf{v} \cdot \Phi^\mu(\mathbf{u},\mathbf{v}) \cdot \mathcal{S}^\mu(\mathbf{u},\mathbf{v})$$

where $\Phi^\mu(\mathbf{u},\mathbf{v}) = \prod_{m,w} \Phi^\mu_{m,w}(\mathbf{u},\mathbf{v})$*, where*

$$\Phi^\mu_{m,w}(\mathbf{u},\mathbf{v}) = \begin{cases} S'_{m,w}(\mathbf{u}_m) \cdot T'_{m,w}(\mathbf{v}_w) & \textit{if } \mu \textit{ matches } m \textit{ and } w \\ 1 - S_{m,w}(\mathbf{u}_m) \cdot T_{m,w}(\mathbf{v}_w) & \textit{otherwise.} \end{cases}$$

$$\mathcal{S}^\mu(\mathbf{u},\mathbf{v}) = \prod_{\substack{m,w \\ m=\mu(w)}} \frac{S_{m,w}(\mathbf{u}_m)}{S'_{m,w}(\mathbf{u}_m)} \cdot \det\left[\frac{S'_{m,w}(\mathbf{u}_m)}{S_{m,w}(\mathbf{u}_m)}\left(1 - \frac{\mathbb{1}[m \neq \mu(w)]}{1 - S_{m,w}(\mathbf{u}_m)T_{m,w}(\mathbf{v}_w)}\right)\right]_{m,w}$$

$$\mathcal{T}^\mu(\mathbf{u},\mathbf{v}) = \prod_{\substack{m,w \\ w=\mu(m)}} \frac{T_{m,w}(\mathbf{v}_w)}{T'_{m,w}(\mathbf{v}_w)} \cdot \det\left[\frac{T'_{m,w}(\mathbf{v}_w)}{T_{m,w}(\mathbf{v}_w)}\left(1 - \frac{\mathbb{1}[w \neq \mu(m)]}{1 - S_{m,w}(\mathbf{u}_m)T_{m,w}(\mathbf{v}_w)}\right)\right]_{m,w}$$

[a] By construction, each integral is properly defined, as ratios in $\mathcal{S}$ and $\mathcal{T}$ will cancel out into polynomial functions. However, we will only be using Theorem 7.9 when $\mathcal{S}$ and $\mathcal{T}$ are defined almost everywhere. A sufficient condition is to ask functions $T_{m,w}$ and $S_{m,w}$ to be strictly decreasing on $\mathbb{R}_+$, for all m, w.

From Theorem 7.9, one can see that a sufficient condition for MPDA and WPDA to output the same distribution is to have $S'_{m,w}(u)/S_{m,w}(u) = T'_{m,w}(v)/T_{m,w}(v)$, for every m, w, u, v. Such condition is satisfied when both $U_{m,w}$ and $V_{m,w}$ are exponentially distributed with mean $P(m, w)$, for all m, w. Intuitively, this corresponds to the case where $U_{m,w}$ and $V_{m,w}$ are Poisson clocks: preference lists are built from the end using the anti-popularity distribution induced by P (see Definition 7.4). Hence, in balanced matching markets with complete preferences, Theorem 7.5 is a corollary of Theorem 7.9.

Lemma 7.10. *Consider a balanced matching market with complete utility preferences, defined by the survival functions $(S_{m,w})$ and $(T_{m,w})$. Then, for every permutation σ,*

$$\mathbb{P}[\sigma \text{ is stable}] = \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} d\mathbf{u} \cdot d\mathbf{v} \cdot \Phi^\sigma(\mathbf{u}, \mathbf{v})$$

where $\Phi^\sigma(\mathbf{u}, \mathbf{v}) = \prod_{m,w} \Phi^\sigma_{m,w}(\mathbf{u}, \mathbf{v})$ and

$$\Phi^\sigma_{m,w}(\mathbf{u}, \mathbf{v}) = \begin{cases} 1 - S_{m,w}(\mathbf{u}_m) \cdot T_{m,w}(\mathbf{v}_w) & \text{if } \sigma(m) \neq w \text{ and } \sigma(w) \neq m \\ -S'_{m,w}(\mathbf{u}_m) \cdot T_{m,w}(\mathbf{v}_w) & \text{if } \sigma(m) \neq w \text{ and } \sigma(w) = m \\ -S_{m,w}(\mathbf{u}_m) \cdot T'_{m,w}(\mathbf{v}_w) & \text{if } \sigma(m) = w \text{ and } \sigma(w) \neq m \\ S'_{m,w}(\mathbf{u}_m) \cdot T'_{m,w}(\mathbf{v}_w) & \text{if } \sigma(m) = w \text{ and } \sigma(w) = m \end{cases}$$

Proof. We proceed as in Section 7.6. Let us define $\mathbf{u} = [U_{m,\sigma^{-1}(m)}]_{m\in\mathcal{M}}$ and $\mathbf{v} = [V_{\sigma^{-1}(w),w}]_{w\in\mathcal{W}}$, the utility each person receives when matched with their predecessor in permutation σ. Conditioning on the values of $\mathbf{u}$ and $\mathbf{v}$, we have the following.

- Each man m such that $\sigma(m) \neq \sigma^{-1}(m)$ prefers $\sigma(m)$ to $\sigma^{-1}(m)$ w.p. $S_{m,\sigma(m)}(\mathbf{u}_m)$.
- Each woman w such that $\sigma(w) \neq \sigma^{-1}(w)$ prefers $\sigma(w)$ to $\sigma^{-1}(w)$ w.p. $T_{\sigma(w),w}(\mathbf{v}_w)$.
- Each pair (m, w) such that $\sigma(m) \neq w$ and $\sigma(w) \neq m$ is blocking w.p. $S_{m,w}(\mathbf{u}_m) \cdot T_{m,w}(\mathbf{v}_w)$.

From Definition 7.8, the probability density functions of $\mathbf{u}_m$ and $\mathbf{v}_w$ are respectively $-S'_{m,\sigma^{-1}(m)}$ and $-T'_{\sigma^{-1}(w),w}$, for all $m \in \mathcal{M}$ and $w \in \mathcal{W}$. Integrating over $\mathbf{u}$ and $\mathbf{v}$ concludes the proof. □

Proof of Theorem 7.9. The first formula, which gives the probability that matching μ is stable, is a corollary of Lemma 7.10. From the second formula, one can deduce the third by symmetry, swapping the roles of men and women. Thus, we now prove the second formula. Combining Lemmas 7.3 and 7.10, we have

$$\begin{aligned} \mathbb{P}[\mu \text{ is stable and men-optimal}] &= \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} d\mathbf{u} \cdot d\mathbf{v} \cdot \sum_{\substack{\sigma \text{ permutation} \\ \sigma_{|\mathcal{W}} = \mu_{|\mathcal{W}}}} (-1)^{C(\sigma)} \cdot \Phi^\sigma(\mathbf{u}, \mathbf{v}) \\ &= \int_{\mathbb{R}_+^{\mathcal{M}}} \int_{\mathbb{R}_+^{\mathcal{W}}} d\mathbf{u} \cdot d\mathbf{v} \cdot \Phi^\mu(\mathbf{u}, \mathbf{v}) \cdot \sum_{\substack{\sigma \text{ permutation} \\ \sigma_{|\mathcal{W}} = \mu_{|\mathcal{W}}}} (-1)^{C(\sigma)} \cdot \frac{\Phi^\sigma(\mathbf{u}, \mathbf{v})}{\Phi^\mu(\mathbf{u}, \mathbf{v})} \end{aligned}$$

Most terms in each product Φ^σ are identical to terms in Φ^μ.

$$\begin{aligned}
\frac{\Phi^\sigma(\mathbf{u},\mathbf{v})}{\Phi^\mu(\mathbf{u},\mathbf{v})} &= \prod_{\substack{m,w\\ \sigma(m)\neq w\\ \mu(m)=w}} \frac{\Phi^\sigma_{m,w}(\mathbf{u},\mathbf{v})}{\Phi^\mu_{m,w}(\mathbf{u},\mathbf{v})} \cdot \prod_{\substack{m,w\\ \sigma(m)=w\\ \mu(m)\neq w}} \frac{\Phi^\sigma_{m,w}(\mathbf{u},\mathbf{v})}{\Phi^\mu_{m,w}(\mathbf{u},\mathbf{v})}\\
&= \prod_{\substack{m,w\\ \sigma(m)\neq w\\ \mu(m)=w}} \frac{S'_{m,w}(\mathbf{u}_m)\cdot T_{m,w}(\mathbf{v}_w)}{S'_{m,w}(\mathbf{u}_m)\cdot T'_{m,w}(\mathbf{v}_w)} \cdot \prod_{\substack{m,w\\ \sigma(m)=w\\ \mu(m)\neq w}} \frac{S_{m,w}(\mathbf{u}_m)\cdot T'_{m,w}(\mathbf{v}_w)}{1-S_{m,w}(\mathbf{u}_m)\cdot T_{m,w}(\mathbf{v}_w)}\\
&= \prod_{\substack{m,w\\ \sigma(m)\neq w\\ \mu(m)=w}} \frac{T_{m,w}(\mathbf{v}_w)}{T'_{m,w}(\mathbf{v}_w)} \prod_{\substack{m,w\\ \sigma(m)=w\\ \mu(m)\neq w}} \underbrace{\frac{T'_{m,w}(\mathbf{v}_w)}{T_{m,w}(\mathbf{v}_w)} \cdot \left(\frac{1}{1-S_{m,w}(\mathbf{u}_m)T_{m,w}(\mathbf{v}_w)} - 1\right)}_{-B_{m,w}(\mathbf{u},\mathbf{v})}
\end{aligned}$$

Let us define a matrix $B(\mathbf{u},\mathbf{v})$, with a diagonal induced by μ.

$$\forall m,w, \qquad B_{m,w}(\mathbf{u},\mathbf{v}) = \frac{T'_{m,w}(\mathbf{v}_w)}{T_{m,w}(\mathbf{v}_w)}\left(1 - \frac{\mathbb{1}[w\neq\mu(m)]}{1-S_{m,w}(\mathbf{u}_m)T_{m,w}(\mathbf{v}_w)}\right)$$

Conveniently, we can rewrite products such that terms with $w=\mu(m)=\sigma(m)$ cancel out.

$$\frac{\Phi^\sigma(\mathbf{u},\mathbf{v})}{\Phi^\mu(\mathbf{u},\mathbf{v})} = (-1)^{|\{m\,|\,\sigma(m)\neq\mu(m)\}|} \prod_{\substack{m,w\\ \mu(m)=w}} \frac{T_{m,w}(\mathbf{v}_w)}{T'_{m,w}(\mathbf{v}_w)} \prod_{\substack{m,w\\ \sigma(m)=w}} B_{m,w}(\mathbf{u},\mathbf{v})$$

Let $D(\sigma) = |\{m\,|\,\sigma(m)\neq\mu(m)\}| - C(\sigma)$ be the discriminant of permutation σ, also defined as N minus the number of cycles in σ. We recognize Leibniz' determinant formula.

$$\sum_{\substack{\sigma \text{ permutation}\\ \sigma_{|\mathcal{W}}=\mu_{|\mathcal{W}}}} (-1)^{C(\sigma)}\cdot\frac{\Phi^\sigma(\mathbf{u},\mathbf{v})}{\Phi^\mu(\mathbf{u},\mathbf{v})} = \prod_{\substack{m,w\\ \mu(m)=w}} \frac{T_{m,w}(\mathbf{v}_w)}{T'_{m,w}(\mathbf{v}_w)} \cdot \underbrace{\sum_{\substack{\sigma \text{ permutation}\\ \sigma_{|\mathcal{W}}=\mu_{|\mathcal{W}}}} (-1)^{D(\sigma)} \prod_{\substack{m,w\\ \sigma(m)=w}} B_{m,w}(\mathbf{u},\mathbf{v})}_{\det(B(\mathbf{u},\mathbf{v}))}$$

Finally, notice that the right hand side of our last equation is exactly equal to $\mathcal{T}^\mu(\mathbf{u},\mathbf{v})$. □

7.8 Complete anti-popularity preferences

In this section we assume that men and women have complete anti-popularity preferences, and we integrate the formula from the previous section to obtain a matrix which gives the probability that two agents will be matched by the deferred acceptance procedure. Notice that because preferences are not symmetric, MPDA and WPDA do not have the same output distribution.

Theorem 7.11. *Assume that men and women have complete anti-popularity preferences, such that man m gives popularity $P_{m,w}$ to woman w, and woman w gives popularity $Q_{m,w}$ to man m. Let $p_{m,w}$ (resp. $q_{m,w}$) the probability that man m and woman w are matched under MPDA (resp. WPDA). Considering P, Q, p and q as $N\times N$ matrices, we have*

$$p = \sum_{X\in\mathfrak{M}} \lambda_X \cdot (X/P)\circ(X/P)^{-T} \qquad \text{and} \qquad q = \sum_{X\in\mathfrak{M}} \lambda_X\cdot(X/Q)\circ(X/Q)^{-T}$$

where $\mathfrak{M}$ is the set of 0-1 matrices containing at least one perfect matching, where $\circ$ and $/$ denote element-wise operations, where A^{-T} denote the inverse transpose of A, and where λ_X

are computed as follow:

$$\forall X \in \mathfrak{M}, \qquad \lambda_X = (-1)^{||X||_1+N} \cdot \frac{\det(X/P)}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\det(X/Q)}{\prod_w \sum_m (X/Q)_{m,w}}$$

First, observe that for any matrix A, the lines/columns of $A \circ A^{-T}$ sum up to 1. Indeed, one can write $A^{-T} = \mathrm{com}(A)/\det(A)$ using the comatrix of A, and Laplace expansion formula shows that $\det(A) = \sum_i A_{i,j} \cdot \mathrm{com}(A)_{i,j}$ for all j.

In the proof, we build λ_X in such a way that $\sum_X \lambda_X = \sum_\mu \mathbb{P}[\mu \text{ is stable and men-optimal}]$, which proves that coefficients sum up to 1. A direct proof of this fact is surprisingly hard to obtain, and was shown by Ilya Bogdanov and Fedor Petrov on MathOverflow (https://mathoverflow.net/questions/360651/sum-over-0-1-matrices).

Proof. We start by writing the survival functions of utilities corresponding to the anti-popularity preferences. We have:

$$\forall t \in \mathbb{R}_+, \qquad S_{m,w}(t) = e^{-tP_{m,w}} \qquad \text{and} \qquad T_{m,w}(t) = e^{-tQ_{m,w}}$$

Let $\mathfrak{M}_\sigma$ be the set of 0-1 matrices such that coefficients $(m, \sigma(m))$ and $(\sigma(w), w)$ are 1's.

$$\mathbb{P}[\sigma \text{ is stable}] = (-1)^{D(\sigma)+C(\sigma)} \sum_{X \in \mathfrak{M}_\sigma} (-1)^{||X||_1+N} \cdot \frac{\prod_w 1/P_{\sigma(w),w}}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\prod_m 1/Q_{m,\sigma(m)}}{\prod_w \sum_m (X/Q)_{m,w}}$$

Where $N + D(\sigma) + C(\sigma) = |\{(m, w) \mid \sigma(m) = w \text{ or } \sigma(w) = m\}|$.

$$\mathbb{P}[\mu \text{ is stable and men-optimal}] = \sum_{\substack{\sigma \text{ permutation} \\ \sigma_{|\mathcal{W}} = \mu_{|\mathcal{W}}}} (-1)^{C(\sigma)} \cdot \mathbb{P}[\sigma \text{ is stable}] \tag{7.1}$$

$$= \sum_{X \in \mathfrak{M}_\mu} (-1)^{||X||_1+N} \sum_{\substack{\sigma \text{ permutation} \\ \sigma_{|\mathcal{W}} = \mu_{|\mathcal{W}}}} (-1)^{D(\sigma)} \cdot \frac{\prod_w 1/P_{\mu(w),w}}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\prod_m (X/Q)_{m,\sigma(m)}}{\prod_w \sum_m (X/Q)_{m,w}} \tag{7.2}$$

$$= \sum_{X \in \mathfrak{M}_\mu} (-1)^{||X||_1+N} \frac{\prod_w 1/P_{\mu(w),w}}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\varepsilon(\mu) \cdot \det(X/Q)}{\prod_w \sum_m (X/Q)_{m,w}} \tag{7.3}$$

Let $\mathfrak{M}$ be the set of 0-1 matrices containing at least one perfect matching.

$$\mathbb{P}[\text{MPDA matches } m \text{ and } w] = \sum_{\substack{\mu \text{ matching} \\ \mu(w)=m}} \mathbb{P}[\mu \text{ is stable and men-optimal}] \tag{7.4}$$

$$= \sum_{X \in \mathfrak{M}} (-1)^{||X||_1+N} \sum_{\substack{\mu \text{ matching} \\ \mu(w)=m}} \frac{\prod_w (X/P)_{\mu(w),w}}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\varepsilon(\mu) \cdot \det(X/Q)}{\prod_w \sum_m (X/Q)_{m,w}} \tag{7.5}$$

$$= \sum_{X \in \mathfrak{M}} (-1)^{||X||_1+N} \cdot \frac{\det(X/P)}{\prod_m \sum_w (X/P)_{m,w}} \cdot \frac{\det(X/Q)}{\prod_w \sum_m (X/Q)_{m,w}} \cdot (X/P)_{m,w} \cdot (X/P)^{-1}_{w,m} \tag{7.6}$$

□

7.9 Simulations

For this chapter, implementations are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/Probability

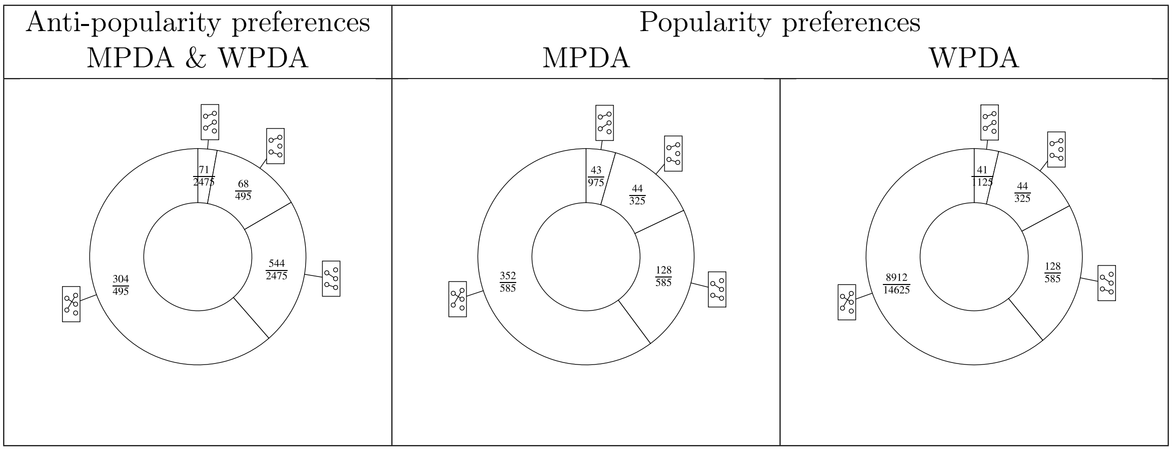


Figure 7.6. Experimental comparison of the output distribution of MPDA and WPDA, under the input distribution described in Figures 7.3, 7.4 and 7.5 with $\omega = 2$.

Output distribution with anti-popularity preferences The output distribution is computed by a C++ program (`main.cpp`), which reads popularities (integers) on the standard input, runs MPDA (class `DeferredAcceptance`), and writes the resulting distribution on the standard output. Preferences of men and women are drawn online: conditioning on each person's preference (variables `stateM` and `stateW`), the program branches each time a man proposes to his next favourite woman (`man.cpp`), and each time a woman answers a proposal (`woman.cpp`). A python script (`run.py`) interacts with the solver and plots the resulting distribution.

Output distribution with popularity preferences When agents have popularity preferences, the conditioning when drawing preferences online is much simpler: a man m will propose to w with probability $P(m,w)/\sum_{w'} P(m,w')$ where the sum is taken over women to whom he has not proposed yet, and woman w will accept with probability $P(m,w)/\sum_{m'} P(m',w)$ where the sum is taken over men who already proposed to w. The output distributions of MPDA and WPDA are computed by a Python program (`popularity.py`)

Match probability with complete popularity preferences In Section 7.8 we assume that agents have complete anti-popularity preferences, and give a formula to compute the probability that a man and a woman will be matched under MPDA or WPDA. Experimentally, a Python program (`matrix.py`) draw random popularity matrices and compare the probability matrix of Theorem 7.11 with the one obtained when running our C++ solver.

7.10 Conclusion and open questions

In this chapter, we study the output distributions of deferred acceptance when agents have anti-popularity preferences: we show that MPDA and WPDA are ex-ante equivalent when preferences are symmetric, and give a close formula for the probabilities of agents being matched. The following questions are left open for future work:

- **Economic interpretation.** The "ex-ante equivalence" property is a mathematical curiosity, which does not imply anything on the strategy-proofness of the deferred acceptance algorithms. However, one economic interpretation is the following. A decision maker who has prior knowledge on the input distribution of preferences (*e.g.* from historical data) might try to favor some outcomes (independently of agents' preferences). We proved that under certain input distributions, a decision maker who has to chose between the MPDA and WPDA procedures cannot manipulate (before seeing agents' preferences). An other potential economic interpretation is that choosing between the two variants of the deferred acceptance does not discriminate towards any community of the market. We leave such an economic study as a very compelling future work.

- **Approximate ex-ante equivalence.** Using a continuity argument, nearly-symmetric anti-popularity preferences should result in approximate ex-ante equivalence. In Section 7.5 we discuss limits for our results, using the more classical model of popularity preferences, and give an example where one matching is arbitrarily more likely to be chosen by one of the two deferred acceptance procedures. Notice this example does not rule out approximate ex-ante equivalence, as the probability for this matching to be chosen by each mechanism is vanishingly small. Simulations from Chapter 8 suggest that approximate ex-ante equivalence should hold when popularities are bounded.

- **Equivalent mechanisms.** Characterizing which other algorithm have the same output distribution as MPDA and WPDA would be an interesting result. Some candidate mechanisms are studied in [KK06]. In particular, numerical simulations suggest it could be the case of the mechanism of *Employment by Lotto* [ACL99] and of Roth and Vande Vate's *incremental procedure* [RV90; Ma96; BCF08].

- **Simplified formula.** In Theorem 7.9, and despite their apparent complexity, formulae for the probability that a matching is stable and men/women optimal should be more tractable than the original probability of stability, in particular because the sum over all matchings should be equal to 1. In a private communication, Joseph Oesterlé and Martin Devaud were able to rewrite the probability as the integral of a differential form, justifying the presence of a determinant as the Jacobian of a differentiable function.

8 | Computing Match Probabilities via Matrix Scaling

This chapter is based on a work in progress, and might lead to a collaboration with authors of [Ash+21].

8.1 Conjecture

In our journey towards answering the question "who gets what?", we derived in Chapter 7 a closed formula to compute the probability that two persons are matched by the deferred acceptance procedure, when agents have complete anti-popularity preferences (Definition 3.3). Unfortunately, our formula (see Theorem 7.11) is not practical from a computational point of view, because of the large number of terms involved in the sum. In this chapter, we attempt to circumvent this computability issue, allowing ourselves to approximate the match probabilities: we will consider matching markets with a large number of agents, and we are interested in the asymptotic match probabilities.

Consider a balanced matching market where agents have complete popularity preferences (see Definition 3.2), where man m gives popularity $P_{m,w}$ to woman w, and woman w gives popularity $Q_{m,w}$ to man m. Denote $p \in \mathbb{R}_+^{\mathcal{M}\times\mathcal{W}}$ (resp. q) the matrix containing the match probabilities in the men-optimal (resp. women-optimal) stable matching. In such markets, stable matchings are perfect, in the sense that everyone is matched, and thus both p and q are doubly stochastic (each line/column sum up to one).

Before stating our conjecture, we examine the men-proposing deferred acceptance procedure as a stochastic process, where men and women draw their preferences online. Each time a man propose, he draws at random his next favorite stable partner; and each time a woman receives a proposal, she tosses a (biased) coin to decide if she accepts it. Observe that man m will propose next to woman w with probability proportional to $P_{m,w}$, and that w will accept with probability (approximately) proportional to $Q_{m,w}$. Thus m and w will be tentatively matched with probability (approximately) proportional to $Z_{m,w} = P_{m,w} \cdot Q_{m,w}$. If the same property holds when the algorithm terminates, then $p_{m,w}$ would be (approximately) proportional to $Z_{m,w}$. Because p is a doubly stochastic matrix, this property is reminiscent of Sinkhorn's Theorem [Sin64] which states that every positive square matrix Z is proportional to a unique doubly stochastic matrix X. Our conjecture states that p is approximately equal to X.

Conjecture 8.1. *Let $C \geq 1$ be a fixed constant. Assume that N men and N women have popularity preferences induced respectively by $P, Q \in [1, C]^{\mathcal{M}\times\mathcal{W}}$. Define the element-wise product $Z = P \circ Q$, and X doubly stochastic using Theorem 8.2.*

Then for every $(m, w) \in \mathcal{M} \times \mathcal{W}$, the men-proposing deferred acceptance procedure match m and w with probability $p_{m,w} = X_{m,w} + o(1/N)$.

Theorem 8.2 (From [Sin64]). *Given $Z \in \mathbb{R}_{+*}^{N\times N}$, there is a unique $X \in \mathbb{R}_{+*}^{N\times N}$ such that:*

- $\sum_j X_{i,j} = 1$ *for each row i; and* $\sum_i X_{i,k} = 1$ *for each column j;*
- *there exist $S \in \mathbb{R}_{+}^{N}$ and $T \in \mathbb{R}_{+}^{N}$ such that $X_{i,j} = S_i \cdot T_j \cdot Z_{i,j}$ for all (i,j).*

Moreover, the sequence of matrices computed by Algorithm 8.1 converges towards X.

For more details on matrix scaling, see the nice survey by Idel [Ide16].

Algorithm 8.1 RAS method

Input: Matrix $Z \in \mathbb{R}_{+*}^{N\times N}$ with positive coefficients.
While Z is far from being doubly stochastic**, do**
  Divide each row i by $\sum_m Z_{i,j}$
  Divide each column j by $\sum_w Z_{i,j}$
Output: matrix Z.

Getting a formal proof of Conjecture 8.1 is challenging for several reasons. We hinted that deferred acceptance match m and w with probability approximately proportional to $Z_{m,w} = P_{m,w} \cdot Q_{m,w}$, to a coefficient S_m which only depends on m, and to a coefficient T_w which only depends on w. First, formally analyzing deferred acceptance as a stochastic process requires us to condition on random draws made so far, on which the values of S_m and T_m will depend. We conjecture that S_m and T_w are approximately independent, which would allow us to write $\mathbb{E}[X_m] \approx Z_{m,w} \cdot \mathbb{E}[S_m] \cdot \mathbb{E}[T_w]$, but this statement requires a formal proof. Second, trying to formalize approximate proportionality, one could write $p_{m,w} = S_m \cdot T_w \cdot Z_{m,w} \cdot (1 + \varepsilon_{m,w})$, for some $\varepsilon_{m,w}$. Thankfully, the function which maps Z to X is continuous, and the decomposition of Z will be close to the decomposition of $Z \circ (1 + \varepsilon)$. Classical analyses of deferred acceptance with random preferences could allow us to show that $|\varepsilon_{m,w}| = \mathcal{O}(1/\ln N)$, but continuity results from the matrix scaling literature requires $|\varepsilon_{m,w}| = \mathcal{O}(1/N)$. Thus, a formal proof would require to close the gap, or to use additional properties such as the approximate independence of $\varepsilon_{m,w}$'s.

Related works. In a recent work, Ashlagi, Braverman, Saberi, Thomas and Zhao [Ash+21] consider large balanced markets where agents have complete aligned popularity preferences, such that each popularity is bounded between 1 and a constant C. Rephrasing their results, they show that every fixed man-woman pair is chosen by the deferred acceptance procedure with probability $\sim 1/N$. The proof follow from a rather technical analysis of deferred acceptance as a stochastic process, similar to the one we sketched above: to compute with whom man m is matched, run deferred acceptance such that m proposes last; if we pause the execution when a proposal from m is accepted for the first time by a woman w, then m and w are likely to stay matched until the end of the algorithm.

Observe that this result is a special case of Conjecture 8.1: assuming agents have aligned preferences, each man m has a popularity P_m, and each woman w has a popularity Q_w; then $S_m = 1/(P_m\sqrt{N})$ and $T_w = 1/(Q_w\sqrt{N})$ yields a valid doubly stochastic decomposition with $X_{m,w} = 1/N$. Both in [Ash+21] and in Conjecture 8.1, the assumption that popularities are bounded between 1 and C is crucial. Indeed, assuming that both man m_i and woman w_i have popularity 2^{-i}, then Theorem 4.19 shows that stable matchings are assortative: a man m_i and a woman w_i such that $|i - j| \gg 1$ will not be matched, with high probability.

Takeaway message. Compared to Theorem 7.11, Conjecture 8.1 gives a computationally efficient but approximate formula for the probability of two persons being matched by the deferred acceptance procedure. Observe that the formula in Conjecture 8.1 is symmetric in P and Q. By swapping matrices P and Q, one effectively computes the match probabilities under the women-proposing deferred acceptance procedure. Comparing this observation to the results of Chapter 7, our conjecture states that the men and women deferred acceptance procedure are asymptotically *ex-ante equivalent* when men and women have popularity preferences (which are not necessarily symmetric).

8.2 Simulations

Implementations of two-sided matching markets with popularity preferences are shared between Chapters 4, 8 and 5, and are available at the following address:

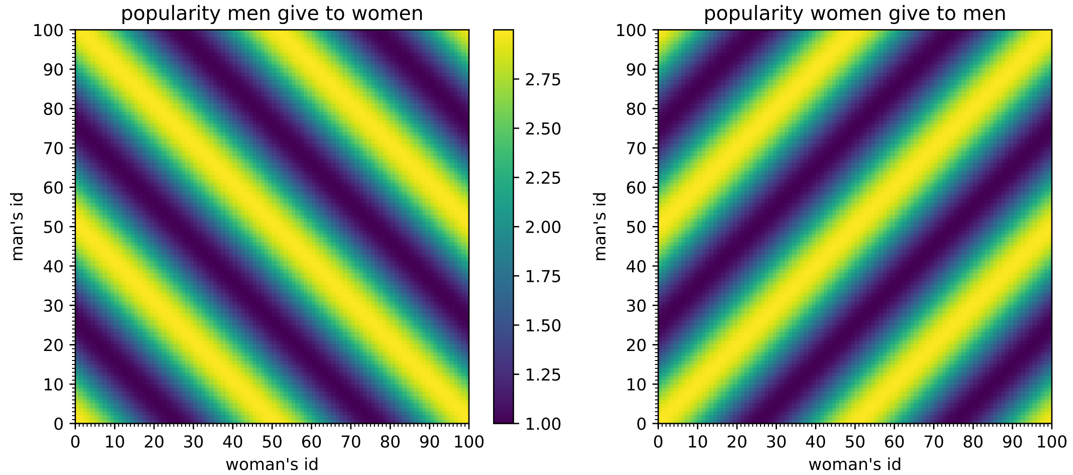
https://github.com/simon-mauras/stable-matchings/tree/master/Popularity

Figure 8.1 illustrates Conjecture 8.1, showing that the scaled products of popularities approximate the match probabilities in the men and women optimal stable matchings. One can observe that with only 100 men and 100 women, the distributions induced by the men-proposing and women-proposing deferred acceptance procedures are different: the men-optimal stable matching gives more importance to the popularities men give to women, and the women-optimal stable matching give more importance to the popularities women give to men.

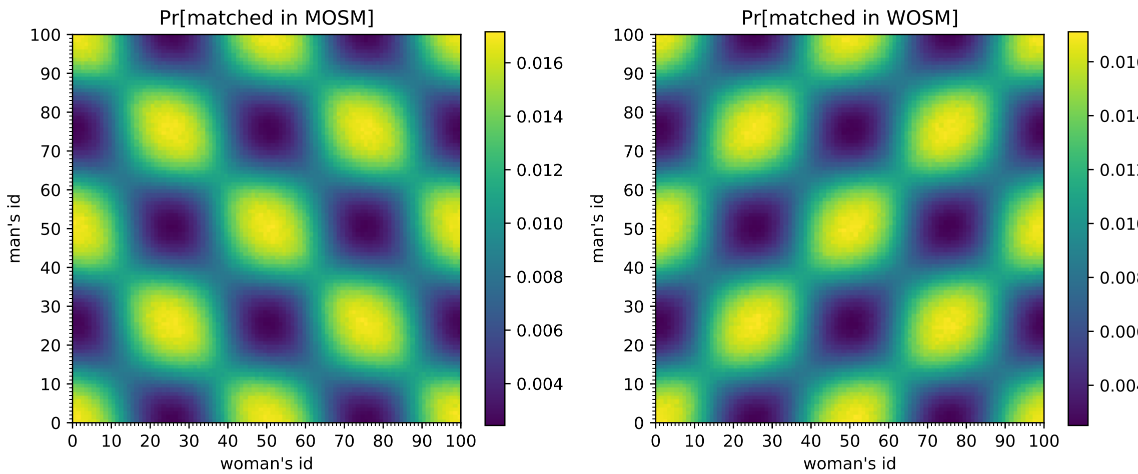
Figure 8.2 illustrates the sensitivity of Conjecture 8.1 to the imbalance of the market. Ashlagi, Kanoria and Leshno [AKL17] show that in matching markets with uniform preferences, adding one woman collapses the different stable matchings into a nearly unique matching, which is close to the original men-optimal one. It is easy to show that the same property holds if agents have bounded popularity preferences. In Figure 8.2, we add one woman from panel (a) to panel (b), and the match probabilities under the women-optimal stable matching become identical to the men-optimal ones. In panel (c), we observe that when adding multiple women, the popularities women give to men play a less and less significant role.

To illustrate the sensitivity of Conjecture 8.1 to the max to min ratio of popularities, we can look at simulations from Chapters 4 and 5. In Figure 4.5, preferences of men and women are aligned, thus the doubly stochastic decomposition yield a matrix X uniformly equal to $1/N$. Theorem 4.19 shows that vertical preferences induce assortative matchings where each agent is matched with someone roughly in front of them. In Figure 5.5, preferences of agents are reversed, in such a way that $Z_{m,w} = 2^{N-1}$ for every m and w, which also yield a uniform doubly stochastic matrix X. We observe that the match probabilities in the men-optimal stable matching are (roughly) proportional to the popularities men give to women. Conversely, the match probabilities in the women-optimal stable matching are (roughly) proportional to the popularities women give to men.

(a) Popularity men give to women and women give to men.



(b) Probability of being matched, in the men and women optimal stable matchings.



(c) Approximating the match probabilities with Conjecture 8.1.

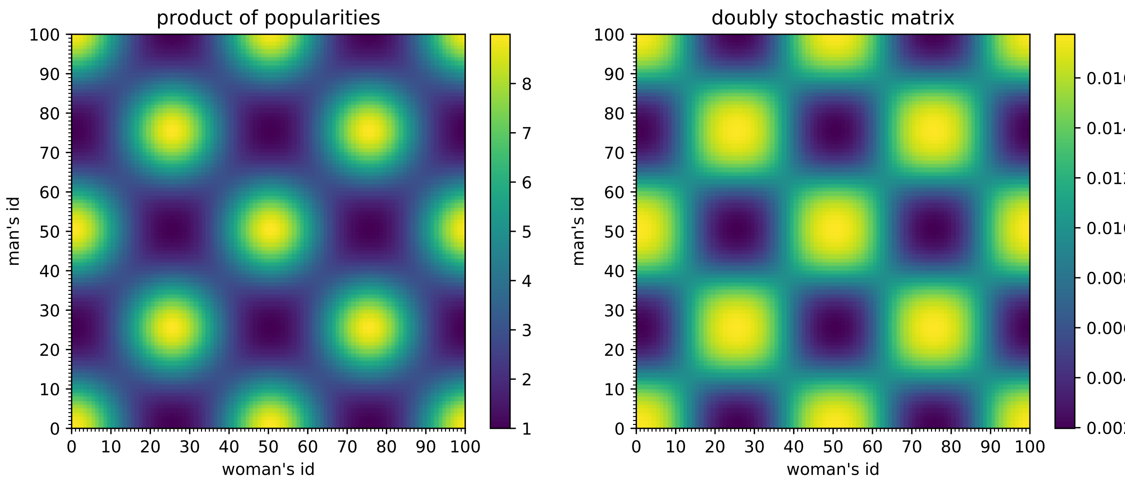
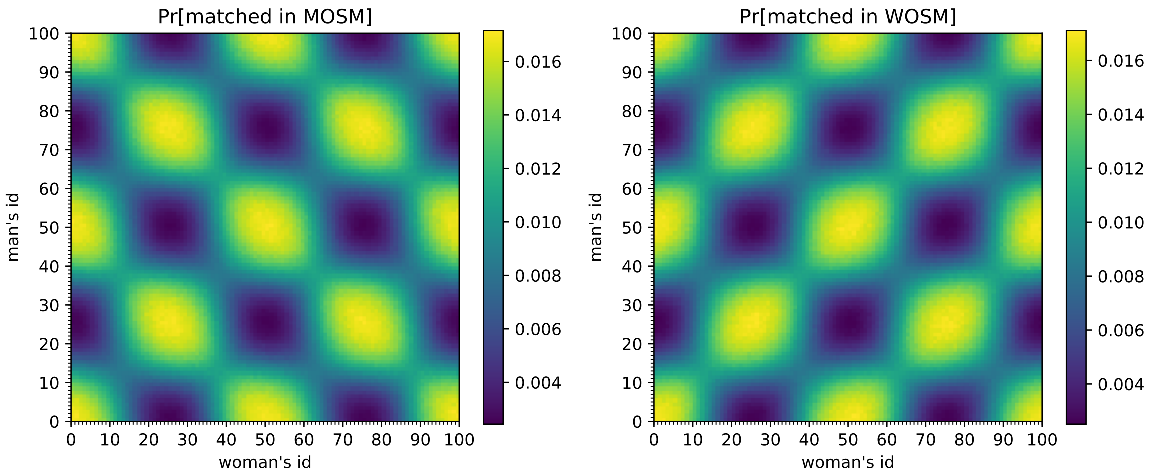
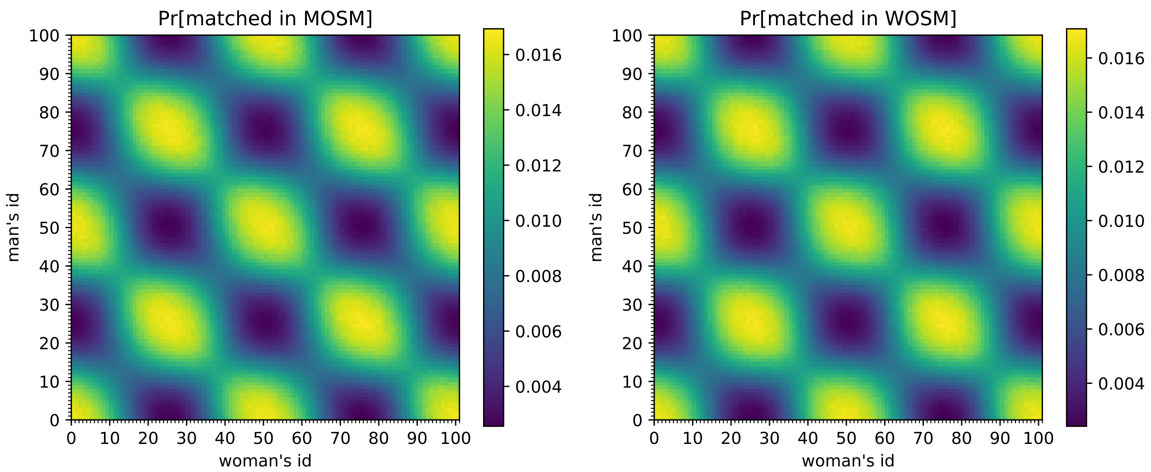


Figure 8.1. Illustrating Conjecture 8.1. The popularity man m_i gives to woman w_j is equal to $P_{m_i,w_j} = 2 + \cos(\frac{(i+j)\pi}{25})$, and the popularity woman w_j gives to man m_i is equal to $Q_{m_i,w_j} = 2 + \cos(\frac{(i-j)\pi}{25})$. Panel (b) contain match probabilities observed over 10^6 runs, and are approximately equal to the doubly stochastic matrix of panel (c), although the men optimal stable matching is closer to the preferences of men (diagonals $i + j = C^{te}$) and the women optimal stable matching is closer to preferences of women (diagonals $i - j = C^{te}$).

(a) $M = 100$ men and $W = 100$ women.



(b) $M = 100$ men and $W = 101$ women.



(c) $M = 100$ men and $W = 110$ women.

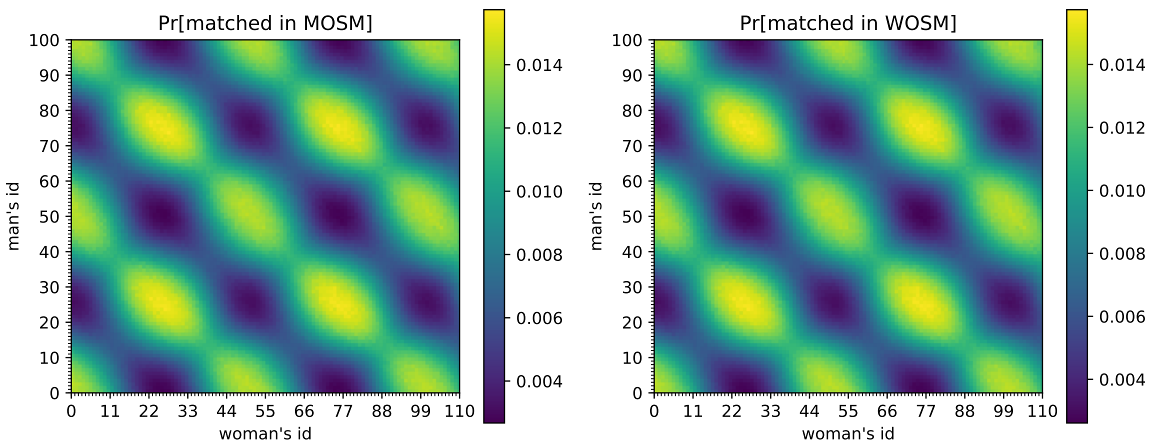


Figure 8.2. Sensitivity of Conjecture 8.1 to the imbalance of the market.

9 | Average Complexity of Daily Deferred Acceptance

9.1 Introduction

In the coupon collector's problem, there are N coupons, and the goal is to compute the expected number of draws required to collect all of them. When k coupons remain, the probability to draw a new one is k/N, and the mean number of draws before a success is N/k. Summing for k from 1 to N gives an expected number of draws of NH_N, where $H_N \sim \ln N$ denotes the harmonic series. Wilson and Knuth [Wil72; Knu76; Knu97] observed that the answer of the coupon collector's problem is an upper bound for the average complexity of deferred acceptance.

The complexity of the men proposing deferred acceptance mechanism is equal to the number of proposals sent by men, which in turn is equal to the sum of rank each man gives to his partner (the size of his preference list if he ends up single). To compute the average complexity, assume that $M = W$ and that men draw their preferences uniformly at random. For the sake of analysis, we assume that men draw their preferences during the execution of the algorithm: this is the principle of deferred decisions. In order to compute an upper bound, we allow men to propose multiple time to the same woman: such proposals will be rejected, and it only increases the total number of proposal. The resulting process is the coupon collector's problem: the sequence of proposals ends when every woman has been drawn at least once.

However, such notion of complexity is not well suited for recent implementations of deferred acceptance. Each year in France, around 800 000 high-school students apply to the centralized college admission procedure. In 2018, the new platform, called Parcoursup, was launched. The main novelty of the procedure is that students do not have to order their applications. Instead, the platform run the school proposing deferred acceptance mechanism, where students answer queries online and have a few days to chose which application they keep each time they receive multiple offers. This mechanism comes with several pros and cons. On the positive side, seats vacated by students leaving the market can be filled quickly by the online procedure; and the fact that students do not have to order applications can decrease self-censorship. On the negative side, the speed of convergence of the procedure becomes of paramount importance, and can be the cause of strategic and non-truthful behaviors from colleges and students.

In this chapter, we analyze the convergence speed of such online mechanisms. Algorithm 9.1 gives an alternative description of the men-proposing deferred acceptance algorithm, where each man can send one offer per day. We looks at the expected number of days required by this procedure, assuming that agents have uniform preferences and that the market is not too imbalanced. We argue that it is crucial that the small side of the market propose: Theorem 9.4 shows that the procedure terminates in a linear number of day when $M \leq W$, and Theorem 9.10 shows that the procedure ends in $\sim W^2/\ln W$ days if $M = W + 1$.

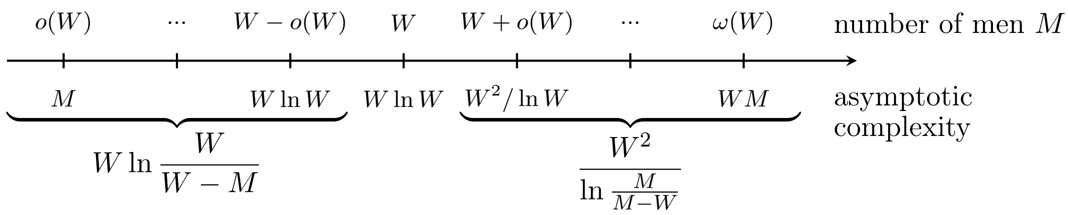


Figure 9.1. Asymptotic sequential complexity of the men-proposing deferred acceptance procedure, when men and women draw their preferences uniformly at random. The formula for $M > W$ differs from [Pit18] by $W(M - W)$, which corresponds to the proposals sent by men who end up single.

Related works. A series of works consider balanced matching markets where men have uniformly random preferences and women have arbitrary preferences. Wilson [Wil72] shows that the coupon collector's problem stochastically dominates the complexity of deferred acceptance, which gives an upper bound of NH_N on the expected complexity. Knuth [Knu76; Knu97] gives an improved upper-bound of $(N-1)H_N + 1$, and conjectures a that the minimum average complexity is reached when women have identical preferences which would give a lower bound of $(N+1)H_N - N$. Knoblauch [Kno07] shows that the lower bound implied by Knuth's conjecture is asymptotically tight, that is the expected complexity is $\sim N \ln N$.

Recent papers study the expected rank each person give to their partner, in variations around the classical balanced uniform model. As discussed in Chapter 4, when the market is unbalanced (even slightly), Ashlagi Kanoria and Leshno [AKL17] show that the set of stable matchings collapses: in every stable matching, agents from the small side are matched with partners of rank $\sim \ln N$, and agents from the large side are matched with partners of rank $\sim N/\ln N$. Subsequent papers from Pittel [Pit18], and Cai and Thomas [CT19] give improved bounds and simplified analyses. As a corollary of their results, the average complexity of the sequential deferred acceptance procedure is a function of the imbalance of the market (see Figure 9.1).

In a college admission setting, each student typically applies to a small number of colleges, and thus each college ranks a small number of students. Kanoria, Min and Qian study the robustness of the results surveyed in Figure 9.1 as a function of the length d of preference lists. They prove the existence of two regimes: if $d = \omega(\ln^2 N)$ then agents from the small (resp. large) side are matched with partners of rank $\sim \ln N$ (resp. $\sim d/\ln N$); and if $d = o(\ln^2 N)$ then all agents are matched with partners of rank $\sim \sqrt{d}$. In particular, this shows that results from this chapter should hold as long as preference lists have size $\omega(\ln^2 N)$.

Takeaway message. In balanced matching markets with uniform preferences, it is easy to see that the duration of the daily deferred acceptance mechanism is at least linear. When all women except one have received offers, then there is exactly one new proposal sent each day. Because the probability that this proposal goes to the last woman is $1/N$, it takes N days in average. But we can go further: in the coupon collector's analysis, when exactly k women did not receive any offers yet, it takes N/k proposals before a successful one, and a pigeonhole principle shows that k new proposals are sent each day, which gives $\approx N/k^2$ days in average. Summing over k gives the series of inverse of squares, which converges towards $\pi^2/6$, and the procedure terminates in $\sim N \cdot \pi^2/6$ days. We show in Section 9.3 that this intuition is tight.

9.2 Daily Deferred Acceptance

In this section, we give a formal description of the Daily Deferred Acceptance mechanism. Each day, men propose to their favorite woman who has not yet rejected them. In particular, a man will keep proposing to the same woman while he is not rejected. Each day, women look at the list of men who proposed to them, and will reject everyone except the best of them. The algorithm stops when each man either reached the end of his list, or is the only person to propose to a woman. For convenience, we will refer to repeated proposals as re-proposals, and new proposals as proposals.

Algorithm 9.1 Daily Men Proposing Deferred Acceptance

Input: Preferences of men $(\succ_m)_{m \in M}$ and of women $(\succ_w)_{w \in \mathcal{W}}$.
Initialization : Start with an empty matching μ.
For $day = 1, 2, \ldots,$ **do**
  Every man (re-)proposes to his favorite woman who has not rejected him yet.
  If every woman received at most one offer that day**, then** break the **for** loop,
  else, women who received multiple offers reject all but the one they prefer.
Output: Resulting matching.

Lemma 9.1. *Algorithm 9.1 outputs the same matching as Algorithm 2.1.*

Proof. Instead of having parallel proposals, we can look at Algorithm 9.1 in such a way that men propose one at a time, skipping over proposals that are identical to ones of previous days. This alternative description is identical to Algorithm 2.1. □

Definition 9.2. In Algorithm 9.1, let X_d be the number of women who have not received any offer the first d days. For all $k \geq 0$, define $T_k = \min\{d \mid X_d \leq k\}$.

Lemma 9.3. *Assume that agents have complete preferences.*

- *If $M \leq W$, then Algorithm 9.1 stops when M women received at least one offer, which corresponds to day T_{W-M}.*
- *If $M > W$, we split the execution of Algorithm 9.1 in two phases: before day T_0 (included) and after day T_0 (excluded). No men reach the end of their list in phase one. The execution stops when exactly $W - M$ men have reached the end of their list.*

Proof. If agents have complete preferences, then a man will never reach the end of his preference list while some woman is still single. If $M \leq W$ and every woman received at most one offer on some day d, then a pigeonhole principle show that exactly $W - M$ women have not received any proposal so far, and thus $d = T_{W-M}$. If $M > W$, no man can reach the end of his list in phase one because some women are still single. The algorithm stops on a day when each woman receive exactly one offer, which occurs when $W - M$ men have reached the end of their list. □

9.3 More Coupons than Collectors

In this section, we discuss the case where there are at most as many men as there are women. Our analysis is based on the fact that Algorithm 9.1 will spend most of its time matching the last k men, which will take approximately $\sum_{i=1}^{k} \frac{W}{i \cdot (W-M+i)}$ days.

Theorem 9.4. *Assume that M men and W women have complete uniform preferences, such that $W - o(\ln W) \leq M \leq W$. In average, the daily deferred acceptance procedure takes $\sim \gamma_{W-M} \cdot W$ days to terminate, where*

$$\forall k \geq 0, \qquad \gamma_k = \sum_{i=1}^{+\infty} \frac{1}{i(i+k)} = \begin{cases} \pi^2/6 & \text{if } k = 0 \\ \frac{1}{k}\sum_{i=1}^{k} \frac{1}{i} & \text{if } k > 0 \end{cases}$$

Proof. We show the upper-bound in Lemma 9.6, and the lower-bound in Lemma 9.9. □

9.3.1 Upper bound

In this section, we show the upper-bound of Theorem 9.4. It holds in a more general case: women can have arbitrary preferences, and M can be much smaller that W.

Lemma 9.5. *Assume that $M \leq W$ and that men draw their preferences uniformly at random. Then, for all $W - M \leq a \leq b \leq W$, we have*

$$\mathbb{E}[T_a - T_b] \leq 1 + \frac{(H_b - H_a) \cdot W}{a + 1 + M - W}$$

where $H_k = \sum_{i=1}^{k} 1/i$ denotes the harmonic series, and $H_0 = 0$ by convention.

Proof. In Algorithm 9.1, we look at the sequential sequence of new proposals made by men, temporarily forgetting about days, and we unfold the classical coupon collector's analysis.

Assuming that i women have not yet received any proposal, the probability that a new proposal is made to one of those women is at least i/W (it is in fact a bit more because the man proposing might already have proposed to some of the other $W - i$ women). Hence, the number of proposal is stochastically dominated by a geometric random variable with a success probability of i/W, whose expected value is W/i. Thus, the expected total number of new proposals made between day T_b and day $T_a - 2$ (we exclude day $T_a - 1$, when the last proposal is made) is $\leq \sum_{i=a+1}^{b} \frac{W}{i}$.

Moreover, on each day up until $T_a - 1$, at least $a + 1$ women have not yet received any proposal, and thus at least $a + 1 + M - W$ men send a new proposal because they were rejected the previous day. Adding one to account for the last day, we obtain

$$\mathbb{E}[T_a - T_b] \leq 1 + \frac{\sum_{i=a+1}^{b} \frac{W}{i}}{a + 1 + M - W} = 1 + \frac{(H_b - H_a) \cdot W}{a + 1 + M - W}$$

□

Lemma 9.6. *Assume that $M \leq W$ and that men draw their preferences uniformly at random. Then $\mathbb{E}[T_{W-M}] \leq \gamma_{W-M} \cdot W + \mathcal{O}(\sqrt{W} \ln W)$.*

Proof. For convenience, we write $\delta = W - M$. Using Lemma 9.3, we know that Algorithm 9.1 stops at day T_δ, and that $T_W = 0$. We are going to bound $\mathbb{E}[T_\delta - T_{\delta+k}]$ and $\mathbb{E}[T_{\delta+k} - T_W]$ for some $k \geq 1$.

$$\mathbb{E}[T_\delta - T_{\delta+k}] = \sum_{i=1}^{k} \mathbb{E}[T_{\delta+i-1} - T_{\delta+i}] \leq k + \sum_{i=1}^{k} \frac{W}{i \cdot (\delta + i)} \leq k + \gamma_\delta \cdot W$$

$$\mathbb{E}[T_{\delta+k} - T_W] \leq 1 + \frac{(H_W - H_{\delta+k}) \cdot W}{\delta + k + 1} \leq 1 + \frac{W \ln W}{\delta + k + 1}$$

We take $k = \lfloor \sqrt{W} \rfloor$, which gives $\mathbb{E}[T_\delta] = \mathbb{E}[T_\delta - T_{\delta+k}] + \mathbb{E}[T_{\delta+k} - T_W] \leq \gamma_\delta \cdot W + \mathcal{O}(\sqrt{W} \ln W)$. □

9.3.2 Lower bound

In the coupon collector's analysis of the sequential complexity, men are amnesiac and are allowed to make redundant proposals, which gives an upper bound on the real number of proposals. To obtain a lower bound, Knuth [Knu76; Knu97] shows that each man makes at most $\mathcal{O}(\ln M)^4/M$ redundant proposals with high probability. We are going to take a similar approach, using a result due to Pittel [Pit89], generalized to unbalanced markets by Ashlagi, Kanoria and Leshno [AKL17].

Lemma 9.7 (Lemma B.4 in [AKL17])**.** *Assume that $M \le W$, and than men and women draw their preferences uniformly at random. Then, in matching $\mu_{\mathcal{M}}$, every man is matched with one of his top $3\ln^2 M$ choices, with probability at least $1 - 1/M^{0.2}$.*

Lemma 9.8. *Assume that $W - o(\ln W) \le M \le W$, and than men and women draw their preferences uniformly at random. Then for all $W - M < k < o(\ln W)$, we have*

$$\mathbb{E}[T_{k-1} - T_k] \ge \left(1 - \frac{\mathcal{O}(1)}{W^{0.1}}\right) \frac{W}{k \cdot (M - W + k)}$$

Proof. If at any point during Algorithm 9.1 a man makes more than $3\ln^2 M$ different proposals, we stop our analysis and use the bound $T_{k-1} - T_k \ge 0$. Using Lemma 9.7, this occurs with probability at most $1/M^{0.2}$. Thus, in the rest of the proof, the probability that a man proposes to a fixed single woman is comprised between $1/W$ and $1/(W - 3\ln^2 M)$.

In Algorithm 9.1, it might be the case that $T_k = T_{k-1}$, because the number of single women jumps from a number $X_d > k$ to a number $X_{d+1} < k$ in one day. Fortunately, this occurs with low probability. Let $d \ge 0$, and condition on the fact that $X_d = K > k$ (exactly K women have not yet received proposals at the end of day d) and that $X_{d+1} < X_d$ (at least one new woman will receive a proposal at day $d+1$). Observe that if $X_{d+1} < k$, then at least $K - k$ other women will receive a new proposal from one of the remaining $M - W + K - 1$ men. Hence

$$\begin{aligned}
\mathbb{P}[X_{d+1} < k \mid X_{d+1} < X_d = K] &\le \frac{(K-1)!/(k-1)!}{(W - 3\ln^2 M)^{K-k}} \binom{M - W + K - 1}{K - k} \\
&\le \frac{(K-1)!/(k-1)!}{(W - 3\ln^2 M)^{K-k}} \left(e\frac{M - W + K - 1}{M - W + k - 1}\right)^{M-W+k-1} \\
&\le \left(\frac{K/e}{W - 3\ln^2 M}\right)^{K-k} \left(\frac{K}{k}\right)^k \left(e\frac{M - W + K - 1}{M - W + k - 1}\right)^{M-W+k-1} \\
&= K^{\mathcal{O}(k)} \exp(-\Omega(K))
\end{aligned}$$

We use the union bound and sum for K from $k+1$ to W. If $k = o(\ln W)$, then the sum is asymptotically dominated by the term $K = k+1$, which gives an upper bound of

$$\mathbb{P}[T_k = T_{k-1}] \le \frac{\exp(\mathcal{O}(k))}{W - 3\ln^2 M} \le \frac{\mathcal{O}(1)}{W^{0.9}}$$

Assuming that $T_{k-1} \ne T_k$, each day between T_k (included) and day T_{k-1} (excluded), exactly k women have not yet received any proposal, and exactly $M - W + k$ men send new proposals because they where rejected the previous day. The probability that some of those men propose to a single woman is

$$\le 1 - \left(1 - \frac{k}{W - 3\ln^2 M}\right)^{M-W+k} \le \frac{k \cdot (M - W + k)}{W - 3\ln^2 M}$$

Thus, we compare $T_{k-1} - T_k$ with a geometric random variable, and obtain

$$\mathbb{E}[T_{k-1} - T_k \,|\, T_{k-1} \neq T_k] \geq \frac{W - 3\ln^2 M}{k \cdot (M - W + k)} = \frac{W}{k \cdot (M - W + k)} \left(1 - \frac{3\ln^2 M}{W}\right).$$

We conclude the proof, combining all sources of error:

$$\left(1 - \frac{3\ln^2 M}{W}\right)\left(1 - \frac{\mathcal{O}(1)}{W^{0.9}}\right)\left(1 - \frac{1}{M^{0.2}}\right) \geq \left(1 - \frac{\mathcal{O}(1)}{W^{0.1}}\right)$$

□

Lemma 9.9. *Assume that $W - o(\ln W) \leq M \leq W$ and that men and women draw their preferences uniformly at random. Then in expectation Algorithm 9.1 finishes after day $W \cdot (\gamma_{W-M} - \mathcal{O}(1/t))$, where $t = \sqrt{(W - M + 1)\ln W}$.*

Proof. For convenience, we write $\delta = W - M = o(\ln W)$ and $t = \lfloor\sqrt{(1+\delta)\ln W}\rfloor = o(\ln W)$. We apply Lemma 9.8 for all $\delta < k < \delta + t$, thus

$$E[T_{W-M}] \geq \left(1 - \frac{\mathcal{O}(1)}{W^{0.1}}\right) \sum_{k=\delta+1}^{\delta+t} \frac{W}{k \cdot (k - \delta)}$$

Re-indexing the sum between 1 and t, we obtain

$$\sum_{k=1}^{t} \frac{W}{k \cdot (k + \delta)} = \sum_{k=1}^{t} \frac{W}{\delta}\left(\frac{1}{k} - \frac{1}{k+\delta}\right) = \frac{W}{\delta}(H_t - H_{t+\delta} + H_\delta) \geq W \cdot (\gamma_\delta - 1/t)$$

To conclude the proof, we hide the error term $\mathcal{O}(1)/W^{0.1}$ inside $\mathcal{O}(1/t)$. □

9.4 More Collectors than Coupons

In this section, we discuss the case where there is more men than women. Our analysis is based on the fact that Algorithm 9.1 will spend most of its time in second phase.

Theorem 9.10. *Assume that $M = N + 1$ men and $W = N$ women have complete uniform preferences. The daily deferred acceptance procedure takes $\sim N^2/\ln N$ days to terminate, in average.*

Proof. Recall that we split the execution of Algorithm 9.1 in two phases: before and after day T_0 (first day when every women received at least one proposal). Recall also that each woman who receives multiple proposals keeps exactly one of them. Thus, a pigeonhole principle shows that for each day of the second phase, exactly one woman receive two proposals (including one which is new), and she rejects one of them. Therefore, in phase two, the number of days is exactly equal to the number of new proposals sent.

In a breakthrough paper, Ashlagi, Kanoria and Leshno [AKL17] show that when $N+1$ men and N women have uniform preferences, the expected number of new proposals is $\sim N^2/\ln N$, both in the sequential Algorithm 2.1 and in the daily Algorithm 9.1. We recall this result in Theorem 9.12, using a tighter bound due to Pittel [Pit18].

In Lemma 9.11, we show that the number of new proposals in the first phase is $\sim N \ln N$. Thus, by linearity of expectation, the number of new proposals in the second phase is $\sim N^2/\ln N$. Because phase two has the same number of days and new proposals, it takes $\sim N^2/\ln N$ days. To conclude the proof, we need to add the number of days of phase one, which is $\sim N$ using Lemma 9.11. □

9.4.1 The first phase is short

In this section, we show that the first phase of Algorithm 9.1 is short, in the sense that both the number of new proposals and the number of days are (quasi-)linear. The analysis is identical to the one of Section 9.3.1

Lemma 9.11. *Assume that $M > W$ and that men draw their preferences uniformly at random. Recall that the first phase of Algorithm 9.1 ends at day T_0. Then,*

- *the expected number of new proposals sent in the first phase is at most $M + W \ln W$,*
- *we have $\mathbb{E}[T_0] \leq \gamma_{M-W} \cdot W + \mathcal{O}(\sqrt{W} \ln W)$, where $\gamma_k = \frac{1}{k}\sum_{i=1}^{k} \frac{1}{i}$.*

Proof. It is relatively easy to bound the total number of new proposals sent. We apply the classical coupon collector's analysis. If k women have not received any proposals yet, the probability that a proposal goes to one of them is at least k/W (it is in fact bigger because the proposer might already have proposed to some women), and by comparison with a geometric variable the expected number of proposals required is at most W/k. Summing for k from 1 to W gives an expected upper bound of $W \ln W$, to which we need to add M to account for the remaining proposals of day T_0.

To compute an upper bound on $\mathbb{E}[T_0]$, we use the exact same analysis as the one of Lemmas 9.5 and 9.6. Between day T_k (included) and day T_{k-1} (excluded), exactly k women have not yet received offers, and exactly $M - W + k$ men send a proposal, thus $\mathbb{E}[T_{k-1} - T_k] \leq 1 + W/(k \cdot (x + M - W))$. We conclude the proof by computing an upper bound on $T_{\lfloor\sqrt{W}\rfloor}$ and summing $\mathbb{E}[T_{k-1} - T_k]$ for k from 1 to $\lfloor\sqrt{W}\rfloor$, . □

9.4.2 Existing results on the sequential complexity

In a breakthrough paper, Ashlagi, Kanoria and Leshno [AKL17] compute the expected rank each person gives to their partner, in unbalanced markets with uniform preferences. Their results can be stated in terms of number of proposals sent in the sequential Algorithm 2.1 and in the daily Algorithm 9.1. We give an asymptotic value in Theorem 9.12, using a tighter bound due to Pittel [Pit18].

Theorem 9.12 (from [Pit18])**.** *Assume that $M = N + 1$ men and $W = N$ women have complete uniform preferences. The total number of proposals sent by men in Algorithm 2.1 is $\sim N^2/\ln N$, in expectation.*

Proof. If we denote P the number of proposals, Pittel shows that for all $0 < a < 1/2$, we have

$$\mathbb{P}\left[\left|\frac{P}{N^2/\ln N} - 1\right| > \frac{1.01}{N^a}\right] \leq \exp(-\theta(N^{1-2a}))$$

Taking $a = 0.4$ shows that $\mathbb{E}[P] = \frac{N^2}{\ln N} \cdot (1 + \mathcal{O}(N^{-0.4}))$. □

9.5 Simulations

For this chapter, implementations are available at the following address:

https://github.com/simon-mauras/stable-matchings/tree/master/DailyGS

In our simulations, a Python script (`run.py`) calls Algorithm 9.1 (`main.cpp`) for different values of M and W. Figure 9.2 plots results for $W = 1000$, averaged over 10^4 realizations.

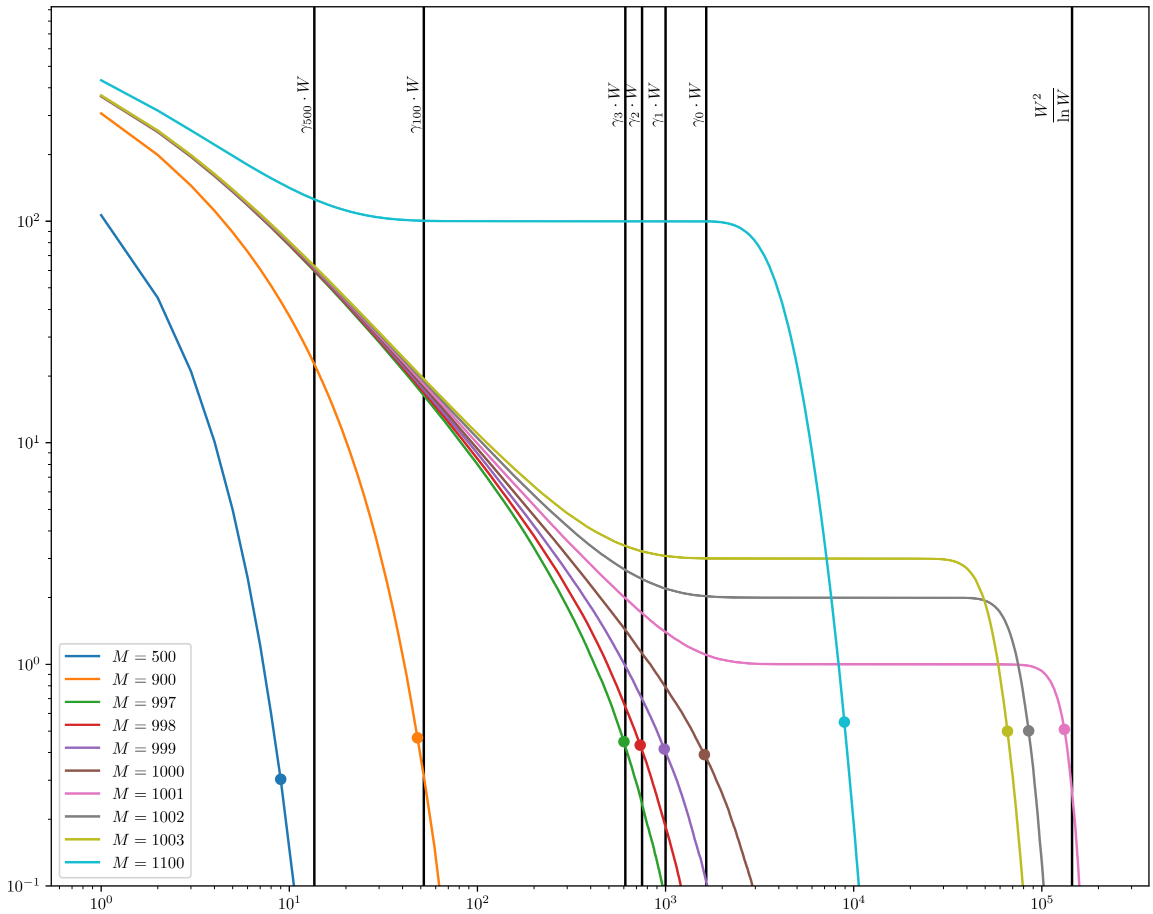


Figure 9.2. Number of days spent by Algorithm 9.1 with $W = 1000$ women. For each number of men M, we plot the number of men rejected each day, averaged over 10^4 realizations. The algorithm stops when exactly 0 men are rejected, because each man either reached the end of his list or was the only one to propose to a woman. The average stopping day is represented by a dot on the plot. Theoretical values from Theorems 9.4 and 9.10 are represented by vertical lines.

More Coupons than Collectors. Section 9.3 discusses the case where $M \leq W$. Theorem 9.4 shows that the expected number of days is asymptotically equal to $\gamma_{W-M} \cdot W$, when $W - M$ is small. Figure 9.2 confirms that our analysis is correct when $W - M \in \{0, 1, 2, 3\}$, and shows that only the upper-bound holds when $W - M \in \{W/10, W/2\}$. In the extreme case where $M = o(\sqrt{W})$, every man proposes to a different women and Algorithm 9.1 terminates in one day. When $M = o(W)$, we have $\gamma_{W-M} \cdot W \sim \ln W$, which shows that our upper-bound is only $\ln W$ away from the correct answer.

More Collectors than Coupons. Section 9.4 discusses the case where $M > W$. More precisely, when $M = W+1$, Theorem 9.10 shows that the number of days is asymptotically equal to $W^2/\ln W$, which is confirmed by Figure 9.2. When $M = W + k$, we will now argue than the number of day must be asymptotically bigger that $W^2/(k \ln W)$. At the end of the first phase, exactly k men are left single. In the second phase, there are k parallel rejection chains, which induce $\sim W^2/\ln W$ new proposals in total. Figure 9.2 illustrates this fact with a plateau where k men are rejected each day. If all k chains were to stop exactly at the same time, the second phase would last $W^2/(k \ln W)$ days. However, chains stop when a man reaches the end of his list, and this might happen earlier/later for some chains.

9.6 Conclusion and open questions

In this chapter, we studied the variant of deferred acceptance where each man can send one proposal per day, and measure the speed in terms of number of days before reaching the men-optimal stable matching. We assumed that men and women have uniformly random preferences, and we adapted the classical coupon collector's analysis. The following questions are left open for future work:

- **Deferred Acceptance with Thresholds.** In the French college admission system, recall that schools send proposals online, and that students who receive multiple offers must report which one they want to keep. One detail on which we did not emphasize is that students can inform the platform when they are not interested by a school, because they already received a better offer. This allows the mechanism to skip proposals, which speeds up the mechanism.

  This situation is equivalent to a version of deferred acceptance with thresholds, where each man knows his rank in the preference list of women, and where women reveal the rank of their best proposal and the end of each day. This way, each man can propose to his favourite woman who would not have rejected him the previous day, skipping over proposals that are doomed to be rejected.

  We leave the analysis of the daily deferred acceptance mechanism with thresholds as a very interesting open question. Notice that in a sequential deferred acceptance mechanism with thresholds, the number of proposals is exactly equal to the number of partner changes by women in Algorithm 2.1. Knuth conjectures that this number should be around $N \ln \ln N$, in expectation (see Problem 1 in [Knu76; Knu97]).

- **Non-uniform preferences.** Extending the results of this chapter when men and women have aligned popularity preferences is an interesting direction. Unfortunately, some technical details complicates the coupon collector's analysis.

  It is always possible to have an upper bound, assuming that men are amnesiac and make redundant proposals, but this upper bound will not be tight if the popularities of coupons are far from being uniform. When popularities have a "tier structure", Ashlagi, Braverman, Saberi, Thomas and Zhao [Ash+21] show that the number of proposals scales with the ratio of the average popularity divided by the minimum popularity.

  As for a lower bound, the most tractable case is when women have identical preferences. If men propose in the corresponding order in Algorithm 2.1, then each man will be matched to the first single woman to whom he proposes, and there is a closed formula for the expected total number of proposals. When men have uniform preferences, Knuth conjectures that women having identical preferences is the best scenario (see Problem 2 in [Knu76; Knu97]) and give an exact formula for the expected number of proposals. We believe that the same conjecture holds when men have popularity preferences.

Bibliography

[Abd+05] Atila Abdulkadiroglu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. "The Boston public school match". In: *American Economic Review* 95.2 (2005), pp. 368–371.

[Abd+20] Atila Abdulkadiroglu, Yeon-Koo Che, Parag A Pathak, Alvin E Roth, and Olivier Tercieux. "Efficiency, justified envy, and incentives in priority-based matching". In: *American Economic Review: Insights* 2.4 (2020), pp. 425–42.

[ACL99] Brian Aldershof, Olivia M Carducci, and David C Lorenc. "Refined inequalities for stable marriage". In: *Constraints* 4.3 (1999), pp. 281–292.

[AKL17] Itai Ashlagi, Yash Kanoria, and Jacob D Leshno. "Unbalanced random matching markets: The stark effect of competition". In: *Journal of Political Economy* 125.1 (2017), pp. 69–98.

[AL16] Eduardo M Azevedo and Jacob D Leshno. "A supply and demand framework for two-sided matching markets". In: *Journal of Political Economy* 124.5 (2016), pp. 1235–1268.

[Alc94] José Alcalde. "Exchange-proofness or divorce-proofness? Stability in one-sided matching markets". In: *Economic design* 1.1 (1994), pp. 275–287.

[APR05] Atila Abdulkadiroglu, Parag A Pathak, and Alvin E Roth. "The new york city high school match". In: *American Economic Review* 95.2 (2005), pp. 364–367.

[AS03] Atila Abdulkadiroglu and Tayfun Sönmez. "School choice: A mechanism design approach". In: *American economic review* 93.3 (2003), pp. 729–747.

[AS21] S Nageeb Ali and Ran I Shorrer. "The College Portfolio Problem". In: (2021).

[AS98] Atila Abdulkadiroglu and Tayfun Sönmez. "Random serial dictatorship and the core from random endowments in house allocation problems". In: *Econometrica* 66.3 (1998), pp. 689–701.

[Ash+21] Itai Ashlagi, Mark Braverman, Amin Saberi, Clayton Thomas, and Geng Zhao. "Tiered Random Matching Markets: Rank Is Proportional to Popularity". In: *12th Innovations in Theoretical Computer Science Conference (ITCS)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2021.

[Ban+13] Abhijit Banerjee, Esther Duflo, Maitreesh Ghatak, and Jeanne Lafortune. "Marry for what? Caste and mate selection in modern India". In: *American Economic Journal: Microeconomics* 5.2 (2013), pp. 33–72.

[BCF08] Péter Biró, Katarína Cechlárová, and Tamás Fleiner. "The dynamics of stable matchings and half-matchings for the stable marriage and roommates problems". In: *International Journal of Game Theory* 36.3-4 (2008), pp. 333–352.

[Bir+20] Péter Biró, Avinatan Hassidim, Assaf Romm, Ran I Shorrer, and Sándor Sóvágó. "Need versus Merit: The Large Core of College Admissions Markets". In: *arXiv preprint arXiv:2010.08631* (2020).

[Bir37] Garrett Birkhoff. "Rings of sets". In: *Duke Mathematical Journal* 3.3 (1937), pp. 443–454.

[BK10] James W Boudreau and Vicki Knoblauch. "Marriage matching and intercorrelation of preferences". In: *Journal of Public Economic Theory* 12.3 (2010), pp. 587–602.

[BST17] Hedyeh Beyhaghi, Daniela Saban, and Eva Tardos. "Effect of selfish choices in deferred acceptance with short lists". In: *arXiv preprint arXiv:1701.00849* (2017).

[BT21] Hedyeh Beyhaghi and Éva Tardos. "Randomness and Fairness in Two-Sided Matching with Limited Interviews". In: *12th Innovations in Theoretical Computer Science Conference (ITCS 2021)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 2021.

[CFV91] Douglas E Critchlow, Michael A Fligner, and Joseph S Verducci. "Probability models on rankings". In: *Journal of mathematical psychology* 35.3 (1991), pp. 294–318.

[Che+04] Shih-Fen Cheng, Daniel M Reeves, Yevgeniy Vorobeychik, and Michael P Wellman. "Notes on equilibria in symmetric games". In: (2004).

[CHK10] Caterina Calsamiglia, Guillaume Haeringer, and Flip Klijn. "Constrained school choice: An experimental study". In: *American Economic Review* 100.4 (2010), pp. 1860–74.

[CK07] Onur B Celik and Vicki Knoblauch. "Marriage matching with correlated preferences". In: (2007).

[Cla06] Simon Clark. "The uniqueness of stable matchings". In: *Contributions in Theoretical Economics* 6.1 (2006).

[Cor+19] Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, and Boris Epstein. "School Choice in Chile". In: *Proceedings of the 2019 ACM Conference on Economics and Computation* (2019), pp. 325–343.

[CS06] Hector Chade and Lones Smith. "Simultaneous search". In: *Econometrica* 74.5 (2006), pp. 1293–1307.

[CT19] Linda Cai and Clayton Thomas. "The short-side advantage in random matching markets". In: *arXiv preprint arXiv:1910.04406* (2019).

[Dan50] HE Daniels. "Rank correlation and population models". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 12.2 (1950), pp. 171–191.

[Dav13] Manlove David. *Algorithmics of matching under preferences*. World Scientific, 2013.

[DF81] Lester E Dubins and David A Freedman. "Machiavelli and the Gale-Shapley algorithm". In: *The American Mathematical Monthly* 88.7 (1981), pp. 485–494.

[DGS87] Gabrielle Demange, David Gale, and Marilda Sotomayor. "A further note on the stable matching problem". In: *Discrete Applied Mathematics* 16.3 (1987), pp. 217–222.

[Dur19] Rick Durrett. *Probability: Theory and Examples*. 5th. Cambridge University Press, 2019. DOI: 10.1017/9781108591034.

[Ech+20] Federico Echenique, Ruy Gonzalez, Alistair Wilson, and Leeat Yariv. "Top of the Batch: Interviews and the Match". In: *arXiv preprint arXiv:2002.05323* (2020).

[EIV22] Federico Echenique, Nicole Immorlica, and Vijay V. Vazirani. *Online and Matching-Based Market Design*. Cambridge University Press (to appear), 2022.

[EM07] Lars Ehlers and Jordi Massó. "Incomplete information and singleton cores in matching markets". In: *Journal of Economic Theory* 136.1 (2007), pp. 587–600.

[GI89] Dan Gusfield and Robert W Irving. *The stable marriage problem: structure and algorithms.* MIT press, 1989.

[GMM19] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Two-Sided Matching Markets with Correlated Random Preferences Have Few Stable Pairs". In: *arXiv preprint arXiv:1904.03890* (2019).

[GMM21a] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Constrained School Choice with Incomplete Information". In: *arXiv preprint arXiv:2109.09089* (2021).

[GMM21b] Hugo Gimbert, Claire Mathieu, and Simon Mauras. "Two-Sided Matching Markets with Strongly Correlated Preferences". In: *Fundamentals of Computation Theory.* Springer. 2021, pp. 3–17.

[GS62] David Gale and Lloyd S Shapley. "College admissions and the stability of marriage". In: *The American Mathematical Monthly* 69.1 (1962), pp. 9–15.

[GS85a] David Gale and Marilda Sotomayor. "Ms. Machiavelli and the stable matching problem". In: *The American Mathematical Monthly* 92.4 (1985), pp. 261–268.

[GS85b] David Gale and Marilda Sotomayor. "Some remarks on the stable matching problem". In: *Discrete Applied Mathematics* 11.3 (1985), pp. 223–232.

[Gus87] Dan Gusfield. "Three fast algorithms for four problems in stable marriage". In: *SIAM Journal on Computing* 16.1 (1987), pp. 111–128.

[HHA10] Gunter J Hitsch, Ali Hortaçsu, and Dan Ariely. "Matching and sorting in online dating". In: *American Economic Review* 100.1 (2010), pp. 130–63.

[HK09] Guillaume Haeringer and Flip Klijn. "Constrained school choice". In: *Journal of Economic theory* 144.5 (2009), pp. 1921–1947.

[HS14] Ron Holzman and Dov Samet. "Matching of like rank and the size of the core in the marriage problem". In: *Games and Economic Behavior* 88 (2014), pp. 277–285.

[Hua06] Chien-Chung Huang. "Cheating by men in the Gale-Shapley stable matching algorithm". In: *European Symposium on Algorithms.* Springer. 2006, pp. 418–431.

[Ide16] Martin Idel. "A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps". In: *arXiv preprint arXiv:1609.06349* (2016).

[IL86] Robert W Irving and Paul Leather. "The complexity of counting stable marriages". In: *SIAM Journal on Computing* 15.3 (1986), pp. 655–667.

[IM15] Nicole Immorlica and Mohammad Mahdian. "Incentives in large random two-sided markets". In: *ACM Transactions on Economics and Computation* 3.3 (2015), p. 14.

[KK06] Bettina Klaus and Flip Klijn. "Procedurally fair and stable matching". In: *Economic Theory* 27.2 (2006), pp. 431–447.

[KMP90] Donald E Knuth, Rajeev Motwani, and Boris Pittel. "Stable husbands". In: *Proceedings of the first annual ACM-SIAM symposium on Discrete algorithms* (1990), pp. 397–404.

[KMQ21] Yash Kanoria, Seungki Min, and Pengyu Qian. "In which matching markets does the short side enjoy an advantage?" In: *Proceedings of the 2021 ACM-SIAM Symposium on Discrete Algorithms (SODA).* SIAM. 2021, pp. 1374–1386.

[Kno07] Vicki Knoblauch. "Marriage matching: a conjecture of Donald Knuth". In: (2007).

[Knu76] Donald Ervin Knuth. *Mariages stables et leurs relations avec d'autres problemes combinatoires: introduction a l'analysis mathematique des algorithmes-*. Les Presses de l'Universite de Montreal, 1976.

[Knu96] Donald E Knuth. "An exact analysis of stable allocation". In: *Journal of Algorithms* 20.2 (1996), pp. 431–442.

[Knu97] Donald Ervin Knuth. *Stable marriage and its relation to other combinatorial problems: An introduction to the mathematical analysis of algorithms*. Vol. 10. American Mathematical Soc., 1997.

[KP09] Fuhito Kojima and Parag A Pathak. "Incentives and stability in large two-sided matching markets". In: *American Economic Review* 99.3 (2009), pp. 608–27.

[Lee16] SangMok Lee. "Incentive compatibility of large centralized matching markets". In: *The Review of Economic Studies* 84.1 (2016), pp. 444–463.

[LP09] Craig Lennon and Boris Pittel. "On the likely number of solutions for the stable marriage problem". In: *Combinatorics, Probability and Computing* 18.3 (2009), pp. 371–421.

[Luc59] R Duncan Luce. "Individual choice behavior: A Theoretical Analysis." In: (1959).

[Luc77] R Duncan Luce. "The choice axiom after twenty years". In: *Journal of mathematical psychology* 15.3 (1977), pp. 215–233.

[LY14] SangMok Lee and Leeat Yariv. "On the efficiency of stable matchings in large markets". In: *Available at SSRN 2464401* (2014).

[Ma96] Jinpeng Ma. "On randomized matching mechanisms". In: *Economic Theory* 8.2 (1996), pp. 377–381.

[Mau20] Simon Mauras. "Two-Sided Random Matching Markets: Ex-Ante Equivalence of the Deferred Acceptance Procedures". In: *Proceedings of the 21st ACM Conference on Economics and Computation*. 2020, pp. 585–597.

[Mer15] Stephan Mertens. "Small random instances of the stable roommates problem". In: *Journal of Statistical Mechanics: Theory and Experiment* 6 (2015).

[MW70] David G McVitie and Leslie B Wilson. "Stable marriage assignment for unequal sets". In: *BIT Numerical Mathematics* 10.3 (1970), pp. 295–309.

[MW71] David G McVitie and Leslie B Wilson. "The stable marriage problem". In: *Communications of the ACM* 14.7 (1971), pp. 486–490.

[MW85] Paul R Milgrom and Robert J Weber. "Distributional strategies for games with incomplete information". In: *Mathematics of operations research* 10.4 (1985), pp. 619–632.

[NNV21] Ndiame Ndiaye, Sergey Norin, and Adrian Vetta. "Descending the Stable Matching Lattice: How Many Strategic Agents Are Required to Turn Pessimality to Optimality?" In: *International Symposium on Algorithmic Game Theory*. Springer. 2021, pp. 281–295.

[Par19] Vilfredo Pareto. *Manuale di economia politica: con una introduzione alla scienza sociale*. Vol. 13. Società editrice libraria, 1919.

[PI94] Boris G Pittel and Robert W Irving. "An upper bound for the solvability probability of a random stable roommates instance". In: *Random Structures & Algorithms* 5.3 (1994), pp. 465–486.

[Pit18] Boris Pittel. “On likely solutions of the stable matching problem with unequal numbers of men and women”. In: *Mathematics of Operations Research* (2018).

[Pit19] Boris Pittel. “On random stable partitions”. In: *International Journal of Game Theory* 48.2 (2019), pp. 433–480.

[Pit89] Boris Pittel. “The average number of stable matchings”. In: *SIAM Journal on Discrete Mathematics* 2.4 (1989), pp. 530–549.

[Pit92] Boris Pittel. “On likely solutions of a stable marriage problem”. In: *The Annals of Applied Probability* (1992), pp. 358–401.

[PS08] Parag A Pathak and Tayfun Sönmez. “Leveling the playing field: Sincere and sophisticated players in the Boston mechanism”. In: *American Economic Review* 98.4 (2008), pp. 1636–52.

[PSV07] Boris Pittel, Larry Shepp, and Eugene Veklerov. “On the number of fixed pairs in a random instance of the stable marriage problem”. In: *SIAM Journal on Discrete Mathematics* 21.4 (2007), pp. 947–958.

[Rom98] Antonio Romero-Medina. “Implementation of stable solutions in a restricted matching market”. In: *Review of Economic Design* 3.2 (1998), pp. 137–147.

[Rot08] Alvin E Roth. “Deferred acceptance algorithms: History, theory, practice, and open questions”. In: *international Journal of game Theory* 36.3-4 (2008), pp. 537–569.

[Rot82] Alvin E Roth. “The economics of matching: Stability and incentives”. In: *Mathematics of operations research* 7.4 (1982), pp. 617–628.

[Rot85] Alvin E Roth. “The college admissions problem is not equivalent to the marriage problem”. In: *Journal of economic Theory* 36.2 (1985), pp. 277–288.

[RP99] Alvin E Roth and Elliott Peranson. “The redesign of the matching market for American physicians: Some engineering aspects of economic design”. In: *American economic review* 89.4 (1999), pp. 748–780.

[RS20] Ross Rheingans-Yoo and Jane Street. *Large random matching markets with localized preference structures can exhibit large cores*. Tech. rep. Mimeo, 2020.

[RS92] Alvin E Roth and Marilda Sotomayor. “Two-sided matching”. In: *Handbook of game theory with economic applications* 1 (1992), pp. 485–541.

[RV90] Alvin E Roth and JH Vande Vate. “Random Paths to Stability in Two-Sided Matching”. In: *Econometrica* 58.6 (1990), pp. 1475–1480.

[Sin64] Richard Sinkhorn. “A relationship between arbitrary positive matrices and doubly stochastic matrices”. In: *The annals of mathematical statistics* 35.2 (1964), pp. 876–879.

[Tan91] Jimmy JM Tan. “A necessary and sufficient condition for the existence of a complete stable matching”. In: *Journal of Algorithms* 12.1 (1991), pp. 154–178.

[Thu27] Louis L Thurstone. “A law of comparative judgment.” In: *Psychological review* 34.4 (1927), p. 273.

[TST01] Chung-Piaw Teo, Jay Sethuraman, and Wee-Peng Tan. “Gale-shapley stable marriage problem revisited: Strategic issues and applications”. In: *Management Science* 47.9 (2001), pp. 1252–1267.

[Var58] Veeravalli S Varadarajan. “On the convergence of sample probability distributions”. In: *Sankhyā: The Indian Journal of Statistics (1933-1960)* 19.1/2 (1958), pp. 23–26.

[Wil72] LB Wilson. "An analysis of the stable marriage assignment algorithm". In: *BIT Numerical Mathematics* 12.4 (1972), pp. 569–575.

[Yan09] Mihalis Yannakakis. "Equilibria, fixed points, and complexity classes". In: *Computer Science Review* 3.2 (2009), pp. 71–85.