Stable marriages

Preference lists

Suppose there are n women $F_1, \ldots, F_n$ and n men $M_1, \ldots, M_n$ such that each person has a prefence list of the persons of the opposite sex, more precisely, there are strict orderings

$$<_{F_1}, \ldots, <_{F_n}, <_{M_1}, \ldots, <_{M_n}$$

on the set $\{1, \ldots, n\}$ such that women F_j prefers man M_k over man M_l if and only if $k <_{F_j} l$ (and similarly for the preferences of men).

Stable marriages

Marriages

Fix a number n and sets $\{F_1, \ldots, F_n\}$ and $\{M_1, \ldots, M_n\}$ of size n.

Then a *marriage* is a bijection π of the set $\{1, \ldots, n\}$

We identify a marriage π with the binary relation

$$H = \{(F_1, M_{\pi(1)}), \ldots, (F_n, M_{\pi(n)})\} .$$

Definition

A marriage H is *unstable* if there are couples (F_i, M_k) and (F_j, M_l) in H such that

$$l <_{F_i} k \quad \text{and} \quad i <_{M_l} j .$$

In this situation, the pair of the couples (F_i, M_k) and (F_j, M_l) is called *dissatisfied*. A marriage is *stable* if it is not unstable.

Question

Is there always a stable marriage?

Stable marriages

Theorem (Marriage theorem)

Let $<_{F_1}, \ldots, <_{F_n}, <_{M_1}, \ldots, <_{M_n}$ be strict orderings on $\{1, \ldots, n\}$. Then there is a stable marriage with respect to these orderings.

Proof.

The proposal algorithm below computes a stable marriage.

Remark

The marriage theorem is false in general if the distinction between women and men is dropped, i.e., if one considers an even number $2n$ of persons where

- each person has a preference list of all the other persons and
- one asks for a partition of the set of persons into sets of size 2 that is stable in a sense similar to the definition above.

The proposal algorithm

Algorithm Proposal (also referred to as proposal algorithm)

Input: Strict orderings $<_{F_1}, \ldots, <_{F_n}, <_{M_1}, \ldots, <_{M_n}$ on $\{1, \ldots, n\}$.

Let $H = \emptyset$.

While there is an unmarried man.

  Let k be mimimum such that M_k is unmarried.

  Let $i = \min_{<_{M_k}} \{j \colon M_k$ has never proposed to F_j before$\}$.

  If F_i is currently unmarried, then let $H = H \cup \{(F_i, M_k)\}$.

  If F_i is currently married to M_l and $M_k <_{F_i} M_l$,

    then let $H = (H \setminus \{(F_i, M_l)\}) \cup \{(F_i, M_k)\}$.

Output: a stable marriage H.

After the least unmarried man M_k has been fixed, he proposes in consecutive iterations of the while loop to all women that have not already rejected or divorced him, in decreasing order of desirability. Eventually M_k is married to the first such woman F_i where F_i is unmarried or is married to a man M_l that in F_i's view is less desirable than M_k (and M_l becomes unmarried again).

The proposal algorithm

Proposition (Verification of the proposal algorithm)

The proposal algorithm terminates and outputs a stable marriage.

Proof.

In each iteration of the while loop, M_k proposes to some woman, i.e., the value i is always defined because

- if there is an unmarried man, then there must also be an unmarried woman, and
- M_k has not already proposed to any unmarried woman, otherwise she had accepted and had then stayed married.

Each man proposes at most once to each woman, thus the while loop is iterated only finitely often.

When the algorithm terminates there is no unmarried man, hence the computed set H is a marriage.

The proposal algorithm

Proof (continued).

Now assume for a proof by contradiction that the computed marriage H is not stable.
Then H contains dissatisfied couples, i.e., contains couples

$$(F_i, M_k) \text{ and } (F_j, M_l) \quad \text{such that} \quad l <_{F_i} k \text{ and } i <_{M_l} j\,.$$

The sequence of partners a woman marries during the execution of the algorithm is strictly increasing in desirability, hence M_l, who in the view of F_i is more desirable than F_i's final partner, is more desirable than all partners of F_i .

Thus M_l never proposes to F_i, otherwise she would marry him.

This contradicts the fact that M_l is married to F_j, hence has proposed to F_j, while M_l surely proposes to F_i before he proposes the first time to F_j. □

The set of all stable marriages

For the following discussion, we fix any instance of the stable marriage problem and consider the set of all stable marriages for this instance.
We define a relation $\preceq_M$ on the set of stable marriages where

$$H \preceq_M H'$$

holds for two stable marriages H and H' if and only if every man has with H either the same wife as with H' or a wife that he prefers to his wife with H' (that is, with H all men do at least as good as with H').

Proposition

The relation $\preceq_M$ is a partial ordering on the set of all stable marriages (i.e., $\preceq_M$ is reflexive, transitive, and antisymmetric).

The set of all stable marriages

Men's best stable marriage

Is there a stable marriage that is a *least* with respect to the $\preceq_M$ relation, i.e., a stable marriage where every man does at least as good as with any other stable marriage?

We say a woman is *possible* for a man, if there is some stable marriage where the man is married to this woman.

Proposition

A stable marriage H is least with respect to the relation $\preceq_M$ if and only if with H every man is married to the most desirable woman among all women who are possible for him.

The set of all stable marriages

The proposal algorithm favors the preferences of men.

Theorem

The output of the proposal algorithm is a stable marriage that is least with respect to the $\preceq_M$ relation.

Proof.

We proof by induction over the proposals made during a run of the proposal algorithm that whenever a woman rejects or divorces a man, then the woman is not possible for this man.

The theorem then follows because the proposals a man makes are chosen without repetition in order of his preference list, hence every man ends up being married to the most desirable woman that did neither reject nor divorce him. By the proposition above, this means that the ouput is least with respect to the $\preceq_M$ relation.

The set of all stable marriages

Proof (continued).

In the induction step, assume that man M_k proposes to woman F_i, who is currently married to M_l.

Case I: M_k is rejected (i.e., F_i prefers M_l to M_k).

There cannot be a stable marriage where M_k is married to F_i.

Assuming otherwise, in such a marriage, M_l could not be married to F_i, while by induction all women more desirable to him than F_i are not even possible for him, so he must be married to a woman F_j who is strictly less desirable to him than F_i.

Consequently, the couples (F_i, M_k) and (F_j, M_l) are dissatisfied, thus contradicting the assumption that the marriage is stable.

The set of all stable marriages

Proof (continued).

Case II: F_i divorces M_l (i.e., F_i prefers M_k to M_l).

There cannot be a stable marriage where M_l is married to F_i.

Assuming otherwise, in such a marriage, M_k could not be married to F_i, while by induction all women more desirable to him than F_i are not even possible, so he must be married to a woman F_j who is strictly less desirable to him than F_i.

Consequently, the couples (F_i, M_l) and (F_j, M_k) are dissatisfied, thus contradicting the assumption that the marriage is stable.

The set of all stable marriages

Remark

The least stable marriage with respect to the relation $\preceq_M$ is the greatest stable marriage with respect to a relation $\preceq_F$ defined like $\preceq_M$ with roles of men and women interchanged.

So the proposal algorithm yields a stable marriage where every man is married to the most preferrable woman possible, whereas every woman is married to the least desirable man possible.

The average number of proposals

Average-case complexity of the proposal algorithm.

How many proposals are made by the proposal algorithm on average on inputs of order n, where the average is taken over all choices of $2n$ strict orderings $<_{F_1}, \ldots, <_{F_n}$, $<_{M_1}, \ldots, <_{M_n}$?

The average number of proposals is the same as the expected number of proposals when we choose the $2n$ strict orderings uniformly at random from the set of all strict orderings on $\{1, \ldots, n\}$.

We will derive an upper bound on the number of proposals, hence it suffices to consider the case where $<_{F_1}, \ldots, <_{F_n}$ are arbitrary but fixed and the strict orderings $<_{M_1}, \ldots, <_{M_n}$ are chosen at random.

The average number of proposals

Theorem (Average-case complexity of the proposal algorithm.)

The average number of proposals made by the proposal algorithm on inputs of order n is at most $\mathrm{O}(n \log n)$.

The theorem is an immediate consequence of the following lemma.

Lemma

Let a nonzero natural number n and strict orderings $<_{F_1}, \ldots, <_{F_n}$ on $\{1, \ldots, n\}$ be given, and let strict orderings $<_{M_1}, \ldots, <_{M_n}$ on $\{1, \ldots, n\}$ be chosen uniformly and independently at random. Then the expected number of proposals made by the proposal algorithm is at most $\mathrm{O}(n \log n)$.

Proof.

The proof of the lemma is done in three steps.

The average number of proposals

Proof (continued).

Step 1 (replacing random inputs by a random choice).

By the Principle of Deferred Decisions, the probabilities for the various possible runs of proposal algorithm and, in particular, the expected number of proposals remains the same if

> instead of choosing $<_{M_1}, \ldots, <_{M_n}$ in advance and always choosing for the next proposal the woman F_i that is most desirable to M_k and has not already been proposed to by M_k,
>
> **$\mathbf{M_k}$ chooses a woman $\mathbf{F_i}$ uniformly at random from all women to whom he has not already proposed.**

Observe that when F_i is determined in an iteration of the proposal algorithm, the restriction of $<_{M_k}$ to the set of all women to whom M_k has not already proposed has not been relevant before, thus by choice of $<_{M_k}$ all possible strict orderings on the latter set of women are equally probable.

The average number of proposals

Proof (continued).

Step 2 (transition to the amnesiac version).

Next consider an *amnesiac version* of the proposal algorithm where in each iteration M_k chooses the next woman to propose to

> not uniformly at random from the set of all women to whom M_k has not proposed before
>
> **but uniformly at random from the set of all women.**

If an unmarried man proposed another time to the same woman, he will be rejected because this woman must have rejected or divorced him in the past and thus is now married to a more desirable man.

Consequently, the amnesiac and the standard version of the proposal algorithm differ only in so far as with the amnesiac version there may occur additional proposals that are rejected anyway.

In summary, the expected number of proposals for the amnesiac version is at least as large as for the standard version.

The average number of proposals

Proof (continued).

Step 3 (anonymizing men).

With the amnesiac version of the proposal algorithm all men can be viewed as behaving the same, hence for each iteration

> instead of asking which man M_k proposes to which woman and whether he is rejected or not,
>
> **we may simply ask whether a married or an unmarried woman is chosen.**

The amnesiac version makes exactly one proposal per iteration and the algorithm terminates when the last unmarried woman is chosen.

Consequently the amnesiac version of the proposal algorithm can be analyzed like the coupon collector's problem with n types of coupons (where choosing an unmarried women corresponds to obtaining a new type of coupon), and the theorem follows. □

Excursus on Harmonic numbers

Definition

Then nth *harmonic number* is $\mathrm{H}_n = 1 + \frac{1}{2} + \cdots + \frac{1}{n}$.

Proposition

For all n we have for the nth harmonic number H_n

$$\ln n \leq \mathrm{H}_n \leq \ln n + 1 \ .$$

Proof.

Let h be the step-like function on the nonnegative reals that attains the value $1/i$ in the interval $[i-1, i)$.

The integral over h from 1 to n is equal to $\mathrm{H}_n - 1$.

Excursus on Harmonic numbers

Proof (continued).

For $x \geq 1$, the function h is bounded from above by $g^+ \colon x \mapsto 1/x$ and from below by $g^- \colon x \mapsto 1/(x+1)$.

The integral over g^+ from 1 to n is $\ln n$.

The integral over g^- from 1 to n is $\ln(n+1) - \ln 2$.

In summary, we have for all $n \geq 1$

$$\ln n - \ln 2 \leq \ln(n+1) - \ln 2 \leq \mathrm{H}_n - 1 \leq \ln n \ .$$ □

Excursus on the geometric distribution

Definition (Geometric distribution)

The *geometric distribution with probability p* is the distribution on the set of natural numbers where

$$\mathrm{Prob}[0] = 0 \quad \text{and} \quad \mathrm{Prob}[i] = (1-p)^{i-1} p \text{ for all } i > 0 \ .$$

Remark

Consider a not necessarily fair coin with probability p for heads. The random variable that is equal to the number of times we have to flip the coin in order to obtain outcome heads at least once has a geometric distribution with probability p.

Excursus on the geometric distribution

Proposition (Geometric distribution)

Let X be a random variable that is geometrically distributed with nonzero probability p. Then the expected values of X is $1/p$.

Proof.

Let $q = 1 - p$, hence $\mathrm{Prob}[X = i] = q^{i-1}p$ for $i \neq 0$.

The expectation of X is $\mathbf{E}[X] = \sum_{i=1}^{\infty} iq^{i-1}p = p\sum_{i=1}^{\infty} iq^{i-1}$.

For any real x where $-1 < x < 1$, we have

$$1 + x + x^2 + \cdots = \frac{1}{1-x} , \quad \text{hence} \quad 1 + 2x + 3x^2 + \cdots = \frac{1}{(1-x)^2}$$

by differentiating both sides of the former equation (recall that for such x the former power series can be differentiated term by term).

Substituting q for x, we obtain $\mathbf{E}[X] = 1/p$. □

Excursus on the coupon collector's problem

The coupon collector's problem

Consider the random experiment where in every round $s = 1, 2, \ldots$ one of n type of coupons is chosen uniformly at random.

Let X be equal to the minimum s such that all n types of coupons have occurred in rounds 1 through s.

The coupon collector's problem asks for the probabilities of the form $\mathrm{Prob}[X = m]$ and for the expectation of X.

Proposition (Expected waiting time of the coupon collector)

In the coupon collector's problem with n types of coupons, the expected number of rounds before all types of coupons have occured at least once is equal to $n\mathrm{H}_n$, hence is at most $\mathrm{O}(n \log n)$.

Excursus on the coupon collector's problem

Proof.

As before, let X be minimum such that all types of coupons occur in round 1 through X.

Let the ith new outcome occur in round s_i (hence $s_1 = 1$, $s_n = X$).

For $i = 1, \ldots, n$, let $I_i = \{s_{i-1} + 1, \ldots, s_i\}$, where we let $s_0 = 0$.

Let $X_i = |I_i| = s_i - s_{i-1}$ and observe $X = X_1 + \cdots + X_n$.

The probability for a new outcome in round $s \in I_i$ is $p_i = \frac{n-i+1}{n}$.

That is, X_i has a geometrical distribution with probability p_i, hence

$$\mathbf{E}[X] = \sum_{i=1}^{n} \mathbf{E}[X_i] = \sum_{i=1}^{n} \frac{1}{p_i} = \sum_{i=1}^{n} \frac{n}{n-i+1} = n\sum_{i=1}^{n} \frac{1}{i} = n\mathrm{H}_n$$

The proposition follows by the upper bound $\ln n + 1$ for the harmonic number H_n. □