The Short-Side Advantage in Random Matching Markets

Linda Cai
tcai@princeton.edu

Clayton Thomas
claytont@princeton.edu

October 14, 2019

Abstract

A recent breakthrough of Ashlagi, Kanoria, and Leshno [AKL17] found that *imbalance* in the number of agents on either side of a random matching market has a profound effect on the expected characteristics of the market. Specifically, across all stable matchings, the "long side" (i.e. the side with a greater number of agents) receives significantly worse matches in expectation than the short side. If matchings are found via the classic one-side proposing deferred acceptance algorithm, this indicates that the difference between the proposing and the receiving side is essentially unimportant compared to the difference between the long and the short side.

We provide new intuition and a new proof for preliminary results in the direction of [AKL17], namely calculating the expected rank that an agent on the long side has for their optimal stable match.

1 Definitions and Preliminaries

We start with the basic definitions. A matching market is a collection $\mathcal{M}$ of "men" and $\mathcal{W}$ of "women", where each man $m \in \mathcal{M}$ has a ranking over women in $\mathcal{W}$, represented as list ordered from most preferred to least preferred, and vice versa. Lists may be partial, and agents included on the list of some $a \in \mathcal{M} \cup \mathcal{W}$ are called the acceptable partners of a. We denote the collection of preference lists for all agents by P, and we say that an agent a prefers x to y if x is ranked before y on the list of a, or if x is an acceptable partner of a but y is not. We also describe the fact that x is an acceptable partner of a by saying that a prefers x to $\emptyset$.

A *matching* is a set of vertex disjoint edges in the bipartite graph $\mathcal{M} \cup \mathcal{W}$, where (m, w) is an edge if and only if m is acceptable to w and w is acceptable to m. We denote a matching as a function $\mu : \mathcal{M} \cup \mathcal{W} \to \mathcal{M} \cup \mathcal{W} \cup \{\emptyset\}$, where $\mu(i)$ is the matched partner of agent i, or $\mu(i) = \emptyset$ if agent i is unmatched.

For a set of preferences P and any matching μ, a man/woman pair (m, w) is called *blocking* if we simultaneously have 1) w prefers m to $\mu(w)$, and 2) m prefers w to $\mu(m)$. A matching μ is *stable* for a set of preferences P if no unmatched man/woman pair is blocking for P. A *pair* (m, w) is called stable for P if $\mu(m) = w$ in *some* stable matching, and m is called a stable partner of w (and vice-versa).

The classic method of finding some stable matching is the one-side-proposing deferred acceptance.

Theorem 1.1. *Algorithm 1 always computes a stable matching μ_0. Moreover, this is the* man-optimal *stable outcome, in that every man is matched in μ_0 to his best stable partner. In particular, the resulting matching is independent of the execution order.*

Algorithm 1 $MPDA$: Men-proposing deferred acceptance

Let $U = \mathcal{M}$ be the set of unmatched men
Let μ be the all empty matching
while $U \neq \emptyset$ and some $m \in U$ has not proposed to every woman on his list **do**
  Pick such a m (in any order)
  m "proposes" to their highest-ranked woman w which they have not yet proposed to
  if w prefers m to $\mu(w)$ **then**
    If $\mu(w) \neq \emptyset$, add $\mu(w)$ to U
    Set $\mu(w) = m$, remove m from U

We also need the following classical fact:

Theorem 1.2 (Rural Hospital Theorem). *For any set of preferences, the set of unmatched agents is the same across every stable outcome.*

2 The Balanced Case

Consider a random matching market P with n men and n women, where each of the n men have uniformly random rankings over the n women, and each of the n women have uniformly random rankings over the n men. Note that each man ranks each woman and vice-versa. We refer to such a random market as *balanced*, in that there are there are the same number of men and women. Let $MPDA(P)$ denote the man-optimal stable outcome realized by running $MPDA$. By the rank that a man m has for a woman w, we mean the index of w on the preference list of m, ranging from 1 to n, and by the average rank that each man gets in a matching μ, we simply mean the sum over all men m of the rank m has for $\mu(m)$, divided by n. We say that a man m receives a partner w better than rank i when w appears in index at most i on m's list.

It is a classical exercise to show that the expectation of the average rank that each man gets in $MPDA(P)$ is $O(\log n)$. The key insight is that, for random balanced matching markets, $MPDA(P)$ essentially behaves like a "coupon collector" random variable. More precisely,

- The sum of the ranks that the men have for their wives in $MPDA$ is exactly the total number of proposals made.
- Define a "coupon collector" random variable Y via the following random process: at each step, a number from $[n]$ is drawn uniformly at random, and Y is the number of steps required for every number in $[n]$ to be drawn at least once.
- For a balanced matching market (or for any where the number of men is at most the number of women) $MPDA$ terminates as soon as n distinct women are proposed to.
- Thus, the total number of proposals made in $MPDA$ is essentially the random variable Y (more precisely, Y statistically dominates the number of proposals made, because in $MPDA$ men will never propose to women who they have already proposed to)

Thus, by analyzing the expectation of the coupon collector random variable Y, we get the following claim:

Proposition 2.1. *Let P be a random matching market with n men and (at least) n women. Then the expected average rank the men have for their wives is $O(\log n)$.*

Of course, the above claim would also hold for women if we used woman-proposing deferred acceptance. On the other hand, in $MPDA$, since the total number of proposals is $O(n \log n)$, on average each women only receives $O(\log n)$ proposals. A woman's rank for her partner in the man-optimal outcome is the minimum rank over all proposals she recieves in $MPDA$. In a random matching market, these rankings are essentially uniformly distributed over $[n]$. Thus women's expected average ranking of their partners should be $\Omega(\frac{n}{\log n})$. Thus, in balanced random maching markets, on average agents recieve dramatically different outcomes (measured by their rank for their partner) when they are on the proposing side verses the recieving side.

3 Unbalanced Markets and the Effect of Competition

We now turn to study "unbalanced" random matching markets, i.e. ones in which one side (say, the women) has strictly more agents than the other side. For the sake of simplicity, for this paper we will assume there are n men and exactly $n+1$ women.

Surprisingly, this very slight imbalance (which one might expect to be negligible as $n \to \infty$) leads to a large qualitative difference in the set of stable outcomes. The authors of [AKL17] show that, for any $\epsilon > 0$ and n large enough, the expected average rank that a woman has for her best stable partner is $(1-\epsilon)n/\log n$. Moreover, they show that, with high probability, the fraction of agents with multiple stable partners goes to 0 as $n \to \infty$. Informally, they conclude that the effect of competition is that the "short side" essentially picks their matches and the "long side" is essentially picked (regardless of who is actually doing the proposing). Contrast this to the balanced case, where the optimal match of women is $O(\log n)$ in expectation, and where it can be shown that the fraction of agents with multiple stable partners goes to 1 as n goes to infinity [Pit89].

The proof in [AKL17] is a careful and detailed probabilistic analysis of a certain algorithm for converting the man-optimal stable matching into the woman-optimal stable matching. While this technique is extremely powerful, and reveals deep properties about the expected behavior of the set of stable matchings, it is a bit unintuitive and hard to absorb.

In this section, we adapt the classical "coupon collector" arguments to prove a preliminary result about the average quality of stable outcomes in unbalanced random matching markets. More specifically, we show that, in the woman-optimal matching for a random market with n men and $n+1$ women, the expected value of the rank of the best stable of partner of any woman is $\Omega\left(\frac{n}{\log n}\right)$. Up to constants, this matches the expected rank the women would receive in man-proposing deferred acceptance.

Given that the stark difference between the balanced and unbalanced case is counterintuitive at first, let's first get some intuition for why we might expect the unbalanced case to be any different at all. Consider *woman-proposing* deferred acceptance with $n+1$ women and n men. In the balanced case, this roughly corresponded to a coupon collector random variable, where every man had to be proposed to once. But now, because some woman must go unmatched, the algorithm will only terminate once *some woman has proposed to every man.* This is a very different random process, and one can imagine it must run for a much longer time, forcing the proposing women to be matched to much worse partners than in the balanced case. Unfortunately, this random process is fairly difficult to analyze[1], so we turn to a coupon-collector type approach instead.

3.1 Expected Rank

The first key to our proof is the following claim, which is based on the concept of "list truncation" as in in [IM05]. Although this claim was essentially given in [IM05], we prove it here for completeness.

[1] For instance, to get a useful analysis, we'd need to keep track of which woman is currently proposing, which men she has already proposed to, and how likely each man is to accept a new proposal.

Claim 3.1. *Fix a woman w^*. For any $i \in [n]$, let P_i denote the set of preferences resulting from w^* after truncating her list at place i (i.e. marking men ranked worse than i "unacceptable"). Then w^* has a stable partner of rank better than i if and only if w^* is matched in $MPDA(P_i)$.*

Proof. ($\Rightarrow$) Suppose for contradiction that w^* has a stable partner of rank better than i, yet w^* is not matched in $MPDA(P_i)$. Let $\mu_0 = MPDA(P)$ and $\mu' = MPDA(P_i)$, and let μ is a stable matching (for preferences P) where w receives a partner better than rank i. Observe that μ must be stable for preferences P_i, because any pair blocking for P_i must be blocking for preferences P. Thus, the set of matched agents in μ would be identical to that of μ', by the rural hospital theorem 1.2. In particular, w^* should be matched in μ', a contradiction.

($\Leftarrow$) Now, suppose w^* is matched in $\mu' = MPDA(P_i)$. We know $\mu'(w^*)$ is ranked by w^* better than i, so it suffices to prove that μ' is stable for preferences P. Certainly, μ' is stable for preferences P_i. Why might a matching, stable for P_i, not be stable for preferences P? The only way is if the blocking pair (m, w) for P is such that $w = w^*$ and w^* truncated m off her preference list. But w^* only accepts proposals from men ranked better than i, and she got a match, so she can't possibly be matched below m. Thus μ' is stable for preferences P.

$\square$

With the above lemma, the proof sketch is as follows:

- By claim 3.1, woman w^*'s rank for her partner in woman-optimal outcome is the best (i.e. minimum) rank i at which she can truncate her list while still being matched in $MPDA(P_i)$.
- Consider running man-proposing deferred acceptance on a random matching market with n men and $n+1$ women. Similar to the balanced random market, we observe that $n < n+1$, so $MPDA$ terminates as soon as n distinct women have accepted a proposal.
- Now imagine w^* **rejects all proposals she receives**. We run $MPDA$ until all women other than w^* recieve a match. The number of proposals again follows a coupon-collector random variable, and we expect $O(n \log n)$ total proposals. In particular, w^* should get $O(\log n)$ total proposals before the algorithm terminates.
- The rank w^* has for each man who proposes to her is (essentially) a uniformly distributed number in $[n]$. Thus, the expected best (minimum) rank of a proposal she receives is $\Omega\left(\frac{n}{\log n}\right)$, which is also the minimum rank where she can truncate her list and still recieve a match. Thus, in expectation she has no stable partners better than this rank.

Below we make this intuition formal.

Theorem 3.2. *In a random matching market with n men and $n+1$ women, for any women w^*, the expected rank of w^*'s best stable partner is $\Omega\left(\frac{n}{\log n}\right)$.*

Proof. Let P be a random matching market for n men and $n+1$ women, however let one of the women w^* submit an empty list of acceptable men (so w^* rejects all men) while everyone else submits uniformly random full length preference lists. Let Y be denote the number of proposals w^* receive before $MPDA(P)$ terminates. In order to bound Y, we define a similar random variable $\overline{Y}$. Consider a "coupon collector with a dud" random process, where in each round a number in $[n+1]$ is drawn uniformly at random, and the process terminates when every number in $[n]$ has been drawn at least once (the number $n+1$ is the "dud coupon" which does not count towards the termination condition). Let $\overline{Y}$ denote the number of times the number $n+1$ is drawn in such a process.

Claim 3.3. *$\overline{Y}$ statistically dominates Y.*

Proof. $MPDA(P)$ can be formally thought of as the following process (which is just running $MPDA$ with the "principle of deferred decisions" used for the men's preferences and with w^* rejecting all proposals):

1. At the beginning of the process, each women w draws a preference list over all men uniformly at random.
2. In each step, an unmatched man m is chosen uniformly at random. Then m draws a woman w uniformly at random from the list of women m hasn't proposed to yet, and m proposes to w.
3. If $w \neq w^*$ and w was previously matched, w rejects her less preferred proposer. If $w = w^*$, w rejects m.
4. The process terminates if all women other than w^* have been proposed to.

Consider a modified deferred acceptance process $MPDA'$, which is identical to the above except item 2 above is replaced by

2′. In each step, an unmatched man m is chosen uniformly at random. m draws a woman w uniformly at random (regardless of whether m has already proposed to w) and propose to w.

Let Y' denote the number of proposals w^* receive before $MPDA'$ terminates. Observe that, ignoring which man is making each proposal, $MPDA'$ is exactly the same as "coupon collector with dud" process described above. Thus, Y' is identically distributed to $\overline{Y}$.

Observe also that the distribution of proposals of $MPDA(P)$ is identical to the distribution of proposals of $MPDA'$, filtering out all "repeated proposals" where a man proposes to a woman for a second (or third, etc.) time. (Conditioned on a proposal not being ignored, each unmatched man is uniformly selected, and so is each woman he has not proposed to). Thus Y is identically distributed to the filtered number of proposal to w^* in process $MPDA'$. For every event in the $MPDA'$ random process, $Y' \geq Y$. Thus, $Y' = \overline{Y}$ statistically dominates Y. □

Claim 3.4. $\overline{Y} \leq 3 \log n$ *with probability at least* $1/2$

Proof. Call the number $n+1$ the "dud", and the numbers in $[n]$ the "coupons" . For $i = 1, \ldots, n$, let t_i be the step where the i^{th} distinct coupon is drawn for the first time. Denote by X_i the number of draws in $(t_{i-1}, t_i]$. Denote by Y_i the number of proposals to the dud in $(t_{i-1}, t_i]$, so that $\overline{Y} = \sum_{i=1}^{n} Y_i$. Observe that in step t_i, a new number in $[n]$ is drawn, thus the dud is never drawn in step t_i. In interval (t_{i-1}, t_i), we know that either an old coupon (that has already been drawn before) is drawn, or the dud is drawn. Thus the dud is drawn with probability $\frac{1}{i}$ in each step in interval (t_{i-1}, t_i). So Y_i can be viewed as the sum over $X_i - 1$ i.i.d Bernoulli random variables Q_{ij} with probability $q_i = \frac{1}{i}$. We thus get

$$\mathbb{E}\left[Y_i\right] = \mathop{\mathbb{E}}_{x_i \sim X_i}\left[\sum_{j=1}^{x_i-1} \mathbb{E}\left[Q_{ij}\right]\right] = \mathbb{E}\left[(X_i - 1) \cdot q_i\right] = q_i \cdot \mathbb{E}\left[X_i - 1\right] = q_i \cdot \left(\frac{1}{p_i} - 1\right) = \frac{1}{n-i+1}.$$

Furthermore, each X_i is a geometric random variable with success probability $p_i = \frac{n-i+1}{n+1}$. Thus, letting H_n denote the nth harmonic number,

$$\mathbb{E}[Y] = \sum_{i=1}^{n} \mathbb{E}[Y_i] = \sum_{i=1}^{n} \frac{1}{n-i+1} = H_n.$$

And by Markov's inequality, we get

$$\mathbb{P}[Y \leq 3\log n] \geq \mathbb{P}[Y \leq 2H_n] \geq 1/2$$

□

Proof of theorem 3.2. Let X be the set of proposals w^* receive in $MPDA$ if she rejects every men. Denote by $\min(X)$ the minimum rank of men in X for w^*. We have

$$\begin{aligned}\mathbb{E}[\min(X)] &\geq \mathbb{P}[|X| \leq 3\log n] \cdot \mathbb{E}\left[\min(X)\middle| |X| \leq 3\log n\right] \\ &\geq \frac{1}{2} \cdot \mathbb{E}\left[\min(X)\middle| |X| = 3\log n\right]\end{aligned}$$

Now w^*'s rank for men in X can be generated as follows: Consider men in X proposing in a sequential order. In each step w^* gives the current proposing men m a rank which has not yet been assigned to a man uniformly at random (i.e. for the i^{th} man, there are $n-i+1$ ranks still available). While the ranks of the men are not exactly i.i.d. uniform random variables on $[n]$, we show next that the expectation of $\min(X)$ behaves almost as if they are.

Assume after the men in X, a new man m proposes to w^*. Denote the probability that m's rank is the minimum out of $|X|+1$ men as $P_{min}(m)$. Since the ranks are each uniform over $[n]$, each rank is equally likely to be the minimum, i.e. $P_{min}(m) = \frac{1}{|X|+1}$. At the same time, we know that

$$\begin{aligned}P_{min}(m) &= \sum_{k\geq 1} \mathbb{P}[\min(X) = k] \cdot \mathbb{P}[\text{rank of } m < \min(X)] \\ &= \sum_{k\geq 1} \mathbb{P}[\min(X) = k] \cdot \frac{k-1}{n-|X|} = \frac{\mathbb{E}[\min(X)]-1}{n-|X|}.\end{aligned}$$

Thus

$$\mathbb{E}[\min(X)] = \frac{n-|X|}{|X|+1} + 1 = \frac{n+1}{|X|+1}.$$

We conclude that

$$\mathbb{E}[\min(X)] \geq \mathbb{E}\left[\min(X)\middle| |X| = 3\log n\right]/2 = \frac{n+1}{2\cdot(3\log n+1)} = \Omega\left(\frac{n}{\log n}\right)$$

Finally, observe that $\min(X)$ is exactly the lowest index i where woman w^* can truncate her list and still be matched in $MPDA(P_i)$. Thus, by claim 3.1, $\min(X)$ is exactly the rank that w^* has for her best stable partner.

□

Acknowledgements. We thank Matt Weinberg for helpful discussions and Mark Braverman for suggesting the proof outline of section 3.1.

References

[AKL17] Itai Ashlagi, Yash Kanoria, and Jacob D. Leshno. Unbalanced random matching markets: The stark effect of competition. *Journal of Political Economy*, 125(1):69 – 98, 2017.

[IM05] Nicole Immorlica and Mohammad Mahdian. Marriage, honesty, and stability. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2005, Vancouver, British Columbia, Canada, January 23-25, 2005*, pages 53–62, 2005.

[Pit89] Boris Pittel. The average number of stable matchings. *SIAM J. Discret. Math.*, 2(4):530–549, November 1989.