The Annals of Applied Probability
1992, Vol. 2, No. 2, 358–401

ON LIKELY SOLUTIONS OF A STABLE MARRIAGE PROBLEM[1]

BY BORIS PITTEL

Ohio State University

To the memory of Mikhail L'vovich Tsetlin

An (n men–n women) stable marriage problem is studied under the assumption that the individual preferences for a marriage partner are uniformly random and mutually independent. We show that the total number of stable matchings (marriages) is at least $(n/\log n)^{1/2}$ with high probability (whp) as $n \to \infty$ and also that the total number of stable marriage partners of each woman (man) is asymptotically normal with mean and variance close to $\log n$. It is proved that in the male (female) optimal stable marriage the largest rank of a wife (husband) is whp of order $\log^2 n$, while the largest rank of a husband (wife) is asymptotic to n. Earlier, we proved that for either of these extreme matchings the total rank is whp close to $n^2/\log n$. Now, we are able to establish a whp existence of an egalitarian marriage for which the total rank is close to $2n^{3/2}$ and the largest rank of a partner is of order $n^{1/2}\log n$. Quite unexpectedly, the stable matchings obey, statistically, a "law of hyperbola": namely, whp the product of the sum of husbands' ranks and the sum of wives' ranks in a stable matching turns out to be asymptotic to n^3, uniformly over all stable marriages. The key elements of the proofs are extensions of the McVitie–Wilson proposal algorithm and of Knuth's integral formula for the probability that a given matching is stable, and also a notion of rotations due to Irving. Methods developed in this paper may, in our opinion, be found useful in probabilistic analysis of other combinatorial algorithms.

1. Introduction and summary. In a group of n men and n women, each person ranks the members of opposite sex as potential marriage partners. A matching (marriage) M between the set of men and the set of women is called stable if there is no pair (m, w) of a man m and a woman w who are not matched but prefer each other to their partners in the matching. Gale and Shapley [10] proved that at least one stable matching always exists. Moreover, they provided an iterative procedure, with a general step being interpreted as a round of proposals of currently free men to women, which always finds a stable marriage. An alternative ("fundamental") algorithm was developed later by McVitie and Wilson [20]. Its work is described by a sequence of individual proposals as opposed to the rounds of simultaneous proposals of men to women in the Gale–Shapley algorithm. However, both versions result

Received June 1990; revised April 1991.

[1]Research supported by NSF Grant DMS-90-02347.

AMS 1980 *subject classifications*. Primary 60C05, 60F99; secondary 05A05, 05C70, 41A60, 41A63.

Key words and phrases. Stable matching, random preferences, ranks, asymptotic behavior, limit theorems.

in the same set of proposals by each man and in the same matching, with the following remarkable property: It matches every man (every woman, if the roles of the sexes are reversed) with the best partner as compared to any other stable marriage. We denote the corresponding stable matchings by $\mathscr{M}_1$ and $\mathscr{M}_2$.

Importantly, McVitie and Wilson found an extension of the fundamental algorithm. Its key element is a "breakmarriage" operation, and it allows (in principle) the determination of the set of all stable matchings once $\mathscr{M}_1$ is found. In particular, there exists a path of stable matchings connecting $\mathscr{M}_1$ and $\mathscr{M}_2$ such that every matching in the path (except $\mathscr{M}_1$) is obtained from the previous matching via an appropriately selected breakmarriage operation [20].

Using an ingenious reduction to a classic urn scheme, Wilson [23] demonstrated that the expected number of proposals in the fundamental algorithm for the random instance of the problem is bounded above by nH_n, $H_n = 1/1 + \cdots + 1/n$. The worst-case upper bound was found later by Itoga [16].

In a book published in 1976, Knuth [17] undertook a systematic study of the combinatorial and probabilistic aspects of the problem. In particular, he found a better upper bound $(n-1)H_n + 1$ and a lower bound $nH_n - O(\log^4 n)$ for the average-case running time. Knuth also demonstrated that, for each n, there is a stable marriage instance of size n with 2^{n-1} stable matchings. Ten years later, Irving and Leather [14] came up with another construction for which, they conjectured, the total number of stable matchings grew even faster. In fact, the authors of [14] hoped that the sequence they found had the maximum exponential growth. Very recently, Knuth [18] partially confirmed the Irving–Leather conjecture by showing rigorously that the solution of a related recurrence obtained in [14] grows as 2.28^n. In striking contrast, it is very easy to find, for each n, an instance with just one stable matching. What are then the expected value and the likely values of S_n, the total number of stable matchings in a random instance of the stable marriage problem?

The question of estimating $E(S_n)$ was posed by Knuth in [17], and he suggested that it might be answered by means of his integral formula for the probability that a given matching is stable. This was done in [22], where we proved that $E(S_n)$ is asymptotic to $e^{-1}n \log n$. We also found an analogous integral formula for the *conditional* distribution of $Q(M)$, the overall rank of wives (as ranked by their husbands) in a matching M, given that M is stable. [By symmetry, the conditional distribution of $R(M)$, the overall rank of husbands in a matching M, is the same.] This formula enabled us to show that with high probability as $n \to \infty$ (whp, in short) in the male optimal stable matching $\mathscr{M}_1$, the ranks $Q(\mathscr{M}_1)$ and $R(\mathscr{M}_1)$ are asymptotic to $n \log n$ and $n^2/\log n$, while symmetrically $Q(\mathscr{M}_2) \sim n^2/\log n$ and $R(\mathscr{M}_2) \sim n \log n$ for the female optimal stable matching $\mathscr{M}_2$. Thus, whp $\mathscr{M}_1$ heavily favors men at the expense of women, and the situation is reversed in $\mathscr{M}_2$. A weak consequence of this result is that whp $S_n \geq 2$.

Can one do better, bearing in mind that $E(S_n)$ is of order $n \log n$? In particular, is it true that whp S_n is unbounded, that is, $P(S_n \geq \omega(n)) \to 1$ for some $\omega(n) \to \infty$? Knuth, Motwani and Pittel [19] found a positive answer to the last question by proving that, for each woman, whp the total number t_n of her distinct partners in all stable matchings lies between $(\frac{1}{2} - \varepsilon)\log n$ and $(1 + \varepsilon)\log n$, $\forall\, \varepsilon > 0$, that is, $S_n \geq (\frac{1}{2} - \varepsilon)\log n$, whp. The proof was based on a careful probabilistic analysis of a simplified version of the McVitie–Wilson (extended) fundamental algorithm, which determines sequentially all stable partners of a particular woman. [A more elaborate algorithm had been proposed earlier by Gusfield [11] for a problem of determination of all stable pairs, i.e., the pairs (m, w) each forming a match in at least one stable matching.]

In this paper, we suggest a different way to analyze the algorithm described in [19]. It allows us to show that, out of the two bounds mentioned previously, the upper bound is sharp. Moreover, we demonstrate that t_n is asymptotic, in distribution, to R_n, the total number of the (left) record values in a uniform random permutation $\omega = (\omega_1, \dots, \omega_n)$ of the set $\{1, \dots, n\}$, $R_n = |\{1 \leq i \leq n: \omega_i = \max_{1 \leq j \leq i} \omega_j\}|$). Specifically, t_n is Gaussian in the limit, with mean and variance both asymptotic to $\log n$. Consequently, the expected number of all stable pairs is asymptotic to $n \log n$ (Section 2, Theorem 2.1, Corollary 2.2).

It is certain, however, that in order to find a significantly better lower bound for S_n, which holds whp, we need a deeper understanding of the likely structure of the set of stable matchings. To this end, we use the techniques of [22] and find an integral (Knuth-type) formula for the *joint* conditional distribution of the pair of ranks $(Q(M), R(M))$, given that M is stable (Section 3, Lemma 3.1).

An asymptotic study of this distribution reveals a surprising phenomenon: For every $\delta \in (0, 1)$, all stable matchings $\mathscr{M}$ satisfy

$$|n^{-3}Q(\mathscr{M})R(\mathscr{M}) - 1| \leq \delta, \tag{1.1}$$

with probability at least $1 - \exp(-c(\delta)n)$, $c(\delta) > 0$. In other words, if each stable matching $\mathscr{M}$ is represented in a two-dimensional plane (u, v) by $(u(\mathscr{M}), v(\mathscr{M})) = (n^{-3/2}Q(\mathscr{M}), n^{-3/2}R(\mathscr{M}))$, then whp all such points fall into an arbitrarily narrow strip which contains the hyperbola $uv = 1$, $u > 0$, $v > 0$ (!) (see Section 4, Theorems 4.1 and 4.1′).

Our previously mentioned results in [22] do indicate that $Q(\mathscr{M})R(\mathscr{M})$ is close to n^3 for both $\mathscr{M} = \mathscr{M}_1$ and $\mathscr{M} = \mathscr{M}_2$. Still, we had not expected that this property might hold whp simultaneously for all other stable matchings as well. Consequently, the minimum total rank of all partners in a stable matching is whp at least $2(1 - \varepsilon)n^{3/2}$, $\forall\, \varepsilon > 0$.

The next steps require a notion of a rotation in a stable matching. This notion was first introduced by Irving [13] in his study of a stable roommates problem, posed by Gale and Shapely [10] and Knuth [17]. A rotation in a stable matching $\mathscr{M}$ is defined as a cyclically ordered sequence of matched pairs (m_i, w_i), $1 \leq i \leq r$, $(m_{r+1}, w_{r+1}) = (m_1, w_1)$, such that each woman w_{i+1} is the best choice for the man m_i among women to whom he prefers his wife w_i, and who prefer m_i to their husbands in $\mathscr{M}$. Elimination of the rotation, that is, pairing m_i with w_{i+1}, $1 \leq i \leq r$, is a very special case of a (successful) breakmarriage operation and, as such, it results in a new stable matching $\mathscr{M}'$ in which no woman gets a worse husband [20]. However, Irving and Leather

[14] were able to prove that the matching $\mathscr{M}_2$ can be generated by a sequence of rotation eliminations, starting from the matching $\mathscr{M}_1$, and that every such chain has the same number of intermediate stable matchings. What is more, every rotation is contained on every maximal chain in the poset of all stable matchings [11]. These results strengthen considerably what was proven by McVitie and Wilson [20] for a broader class of all breakmarriage operations. (Unlike a rotation, in a general breakmarriage operation a woman may change a suitor more than once.) For an excellent (up-to-date) exposition of these and other related properties of the stable marriage problem and various applications, the reader is referred to a recent book by Gusfield and Irving [12].

Let (M, M') denote a pair of matchings such that M' is obtained from M by breaking up some pairs $(m_1, w_1), \dots, (m_r, w_r)$ in M and pairing m_i with w_{i+1}, $w_{r+1} = w_1$. In Section 3 (Lemma 3.2), we derive two integral formulas. The first is for the probability P_{nr} of the event A that M is stable and that (m_i, w_i), $1 \le i \le r$, form a rotation in M, so that M' is also stable. The second formula is for the conditional distribution of $(Q(M'), R(M))$, given the event A.

We use these results in Section 5 to study the likely behavior of rotations. The formula for P_{nr} enables us to conclude that whp no rotation may have length exceeding $(n \log n)^{1/2}$ (Theorem 5.1). The second formula yields that whp for all pairs $(\mathscr{M}, \mathscr{M}')$ of stable matchings, with $\mathscr{M}'$ obtained from $\mathscr{M}$ through elimination of a rotation in $\mathscr{M}$,

$$|n^{-3} Q(\mathscr{M}') R(\mathscr{M}) - 1| \le \varepsilon, \tag{1.2}$$

$\varepsilon > 0$ (Theorem 5.2). From (1.1) and (1.2) we show then that for each $\varepsilon > 0$, whp for *every* point (u, v) on the hyperbola $uv = 1$, such that $u, v = o(n^{1/4})$, there exists a stable marriage $\mathscr{M}$ for which the distance between $(u(\mathscr{M}), v(\mathscr{M}))$ and (u, v) is at most ε (Theorem 5.3). In particular (picking $u = v = 1$), the minimum total rank of all partners in a stable marriage is whp asymptotic to $2n^{3/2}$. In contrast, for both $\mathscr{M}_1$ and $\mathscr{M}_2$, the total rank is whp asymptotic to $n^2/\log n$.

We should note here that, answering a question posed by Knuth [17], Irving, Leather and Gusfield [15] found an efficient algorithm which determines a minimum total rank (egalitarian) stable marriage in $O(n^4)$ time. Later, Feder [7] proposed an algorithm with a running time $O(n^3 \log n)$. Very recently, he discovered an even faster algorithm, with a time bound of $O(n^3)$ [8]. Our previously mentioned result indicates that the time spent is almost always well worth it, since in the resulting stable marriage $\mathscr{M}_3$ the total rank is whp reduced by a factor close to $n^{1/2}/2 \log n$, as compared with $\mathscr{M}_1$ and $\mathscr{M}_2$.

To continue, let $\mathscr{Q}(\mathscr{M})$ and $\mathscr{R}(\mathscr{M})$ denote the largest rank of a wife and the largest rank of a husband in a stable matching $\mathscr{M}$. We show (Section 6, Theorems 6.1 and 6.2) that whp $\mathscr{Q}(\mathscr{M}_1)$ lies between $(1 - \varepsilon)\log^2 n$ and $(2 + \varepsilon)\log^2 n$, while $\mathscr{R}(\mathscr{M}_1)$ is asymptotic to n. The reversed situation takes place for the female optimal stable matching $\mathscr{M}_2$. On the other hand, whp both $\mathscr{Q}(\mathscr{M}_3)$ and $\mathscr{R}(\mathscr{M}_3)$ lie between $(1 - \varepsilon)n^{1/2} \log n$ and $(\frac{7}{4} + \varepsilon)n^{1/2} \log n$ (Theorem 6.3). So, the largest rank of a marriage partner in $\mathscr{M}_3$ is *at least* about

$n^{1/2}/1.75 \log n$ times smaller than those in $\mathscr{M}_1$ and $\mathscr{M}_2$. What is more, $\min\{\max(\mathscr{Q}(\mathscr{M}), \mathscr{R}(\mathscr{M}))|\mathscr{M}\}$ is whp at least $(1-\varepsilon)n^{1/2} \log n$, too. So, the largest rank of a spouse in the egalitarian marriage $\mathscr{M}_3$ is of the lowest order, in comparison with all other stable marriages, for almost all instances of the problem. We add that the *minimum-regret* stable matching, on which the minimax is achieved, was studied by Knuth [17], who described an $O(n^4)$ time solution (attributed to S. Selkow). Gusfield [11] found an $O(n^2)$ time algorithm for this problem (see also [12]). At the present time it is not clear to us whether the minimum-regret stable matching has the total rank close whp to the total rank of $\mathscr{M}_3$. If it has, that means that Gusfield's algorithm provides whp a near-optimal solution of the egalitarian stable marriage problem.

The preceding assertion regarding $\mathscr{Q}(\mathscr{M}_1)$ [and $\mathscr{R}(\mathscr{M}_2)$] has some interesting consequences, which are discussed in Section 7. First, it implies a lower bound $nH_n - c\log^3 n$ for the expected running time of the fundamental algorithm (Corollary 7.1), which is an improvement of Knuth's lower bound [17]. Second, it means that in this algorithm whp no man makes more than $(2+\varepsilon)\log^2 n$ proposals, [indeed, $(1+\varepsilon)\log^2 n$ proposals; see the Note added in proof at the end of the text], but there is a man who makes at least $(1-\varepsilon)\log^2 n$ proposals. We use this result to show that if a stable marriage problem is chosen at random from the pool of problems in which each man knows exactly k $(\le n)$ women, then a stable marriage exists whp if $k \ge (2+\varepsilon)\log^2 n$ [$k \ge (1+\varepsilon)\log^2 n$] and whp there is no stable marriage if $k \le (1-\varepsilon)\log^2 n$ (Corollary 7.3). [It is worth noticing here that, for a complete matching problem on a random bipartite graph, the threshold value of k is $\log n$ (Erdös and Rényi [6]), thus considerably smaller than for the stable marriage problem.]

Most interestingly, a combination of the whp upper bounds for $\mathscr{Q}(\mathscr{M}_1)$ and the maximum length of a rotation will allow us to prove, via the Irving–Leather theorem [14], that whp S_n (the total number of stable matchings) is at least $(1-\varepsilon)(n/\log n)^{1/2}$, $\forall\, \varepsilon > 0$ (Corollary 7.2). This bound strengthens considerably the logarithmic bound $(\frac{1}{2}-\varepsilon)\log n$ in [19] and its improvement $(1-\varepsilon)\log n$ in this paper, which are both based on estimates of the total number of stable husbands (wives) of a particular woman (man).

In view of this polynomial bound and our earlier estimate $E(S_n) \sim e^{-1} n \log n$, it does not seem too implausible now that, in fact, S_n is whp quasilinear, of order $n/\log n$, say, or even $n \log n$, just like its expectation. For further progress in this direction, much more would have to be learned about likely solutions for the random instance of the stable marriage problem.

2. Stable husbands. Let an instance of the stable marriage problem be given. For a particular woman w_1, introduce $\mathscr{T}(w_1)$, the set of all stable husbands of w_1, that is, $m \in \mathscr{T}(w_1)$ iff m and w_1 are partners in at least one stable matching. The set $\mathscr{T}(w_1)$ can be determined via a *sequential* proposal algorithm described by Knuth, Motwani and Pittel [19] (cf. Gusfield [11]).

Its first phase is the McVitie–Wilson algorithm [20] which produces the stable matching $\mathscr{M}_1$, simultaneously male-optimal and female-pessimal. It works as follows: The (arbitrarily ordered) men propose in turn to women, and each man always proposes to a woman of his first choice among the women who have not rejected him yet. When the $(k+1)$th man makes his first proposal, the first k men are temporarily matched with some k women. If his choice is not one of these women, he is temporarily accepted. Otherwise, his proposal triggers a sequence of collisions, each resolved by a woman in favor of a better suitor at the expense of a worse suitor, who has to propose to his next best choice. This sequence ends when a new woman receives a proprosal; she accepts it and the next $(k+2)$th man may propose. The algorithm stops once all the women have been proposed to, and we obtain the desired matching $\mathscr{M}_1$. Keeping in mind the way the collisions are resolved and also the fact that $\mathscr{M}_1$ is female-pessimal, we see that no woman has been proposed to by a stable husband other than her (worst) stable husband in $\mathscr{M}_1$.

The second phase beings with woman w_1 rejecting her (stable) husband in $\mathscr{M}_1$. He makes a proposal to a woman of his next best choice; in a resulting collision she chooses a better suitor, and a rejected man proposes to his next best choice, and so on. Clearly, we have a matching whenever w_1 receives a proposal. It is postulated that w_1 rejects all the proposals she receives, thereby sustaining the sequence of proposals and rejections for as long as possible. The process terminates when a man is rejected by a woman of his last (nth) choice. It was proven in [19] that all the stable husbands of w_1, besides her husband in $\mathscr{M}_1$, are among those men who propose to w_1 in the second phase. Moreover, a man who proposes to w_1 is her stable husband if and only if w_1 prefers him to all the previous proponents, thus to all the stable husbands determined so far.

Consider an instance of the stable marriage problem in which the ranking of men by women and women by men is chosen uniformly at random from the pool of all $(n!)^{2n}$ possible ranking systems. This assumption is equivalent to a condition that, for each person, every ranking of members of the opposite sex has the same probability $1/n!$ and that the rankings by different members are independent. By symmetry, the distribution of the random variable $t_n(w_1) = |\mathscr{T}(w_1)|$ does not depend on w_1, and also it coincides with the distribution of the total number of stable wives of a man.

Using the preceding algorithm, we shall prove the following statement.

THEOREM 2.1. *$t_n^* =_{\text{def}} (t_n(w_1) - \log n)/(\log n)^{1/2}$ converges in distribution to a normal random variable with zero mean and unit variance; in short, $t_n^* \Rightarrow \mathscr{N}(0,1)$.*

PROOF. Our argument uses a so-called principle of deferred decisions ([17], [19]). It is postulated that the random ranking system is not given in advance, but rather unfolds step by step to the extent necessary for a full run of the proposal algorithm. At each step a man, whose turn it is to propose, proposes to a woman chosen uniformly at random among women he has not proposed to so far. If it is the kth proposal to this woman then the man's rank (relative to the group of k suitors) is distributed uniformly at random on the set $\{1, \dots, k\}$; in particular, his proposal will be the best so far with probability $1/k$. The resulting sequence of proposals and rejections will have exactly the same distribution as the original sequence.

Let $\mathcal{U}_n$ denote the rank of the best stable husband of the woman w_1. Introduce V_{n1} and V_n, the total number of proposals made to w_1 during the first phase of the algorithm and overall, respectively. By symmetry, $\mathcal{U}_n$ coincides in distribution with the rank of the best wife of a man, and this rank equals the number of proposals made by him in the first phase. According to Wilson [23], the expected number of all proposals in this phase is at most nH_n. Therefore,

$$E(\mathcal{U}_n) = E(V_{n1}) \le n^{-1}(nH_n) \\ = \log n + O(1). \tag{2.1}$$

Furthermore, conditioned on $V_n = v$, $\mathcal{U}_n$ coincides in distribution with 1 + [the occupancy number of a fixed cell in the uniform allocation scheme with $(v+1)$ cells and $(n-v)$ *indistinguishable* balls], that is

$$P(\mathcal{U}_n > u | V_n = v) = \binom{n-u}{v} \Big/ \binom{n}{v}, \qquad u + v \le n. \tag{2.2}$$

A simple estimate shows then that

$$P(\mathcal{U}_n > u | V_n = v) \ge 1 - uv/(n-v). \tag{2.3}$$

Let $\omega(n) \to \infty$, however slowly. We want to prove that

$$\lim_{n\to\infty} P(V_n > v_n) = 1, \quad \text{if } v_n = \frac{n}{\omega(n)\log n}. \tag{2.4}$$

To this end, we write

$$P(V_n \le v_n) \le P(\mathcal{U}_n > u_n) + P(\mathcal{U}_n \le u_n, V_n \le v_n),$$

where

$$u_n = \omega(n)^{1/2} \log n.$$

Here, by (2.1) and Markov's inequality,

$$P(\mathcal{U}_n > u_n) \le u_n^{-1} E(\mathcal{U}_n) = O\big(\omega(n)^{-1/2}\big).$$

Also, using (2.3),

$$\begin{aligned} P(\mathcal{U}_n \le u_n, V_n \le v_n) &= \sum_{v \le v_n} P(\mathcal{U}_n \le u_n | V_n = v) P(V_n = v) \\ &\le \frac{u_n v_n}{n - v_n} \sum_{v \le v_n} P(V_n = v) \\ &= O\big(\omega(n)^{-1/2}\big). \end{aligned}$$

So, $P(V_n \le v_n) \to 0$ and relation (2.4) follows.

Now, $t_n(w_1) = B_n - B_{n1}$, where B_n and B_{n1} are the total number of the currently best proposals in the sequence of all V_n proposals and in the sequence of the first $V_{n1} - 1$ proposals, respectively. Conditioned on $V_n = v$

($V_{n1} = v$, respectively), the random variable B_n (B_{n1}, respectively) has the same distribution as R_v, the total number of the (left) record values in a uniform random permutation $\omega = (\omega_1, \dots, \omega_v)$ of the set $\{1, \dots, v\}$, that is,

$$R_v = \left| \left\{ 1 \le i \le v \colon \omega_i = \max_{1 \le j \le i} \omega_j \right\} \right|.$$

Since

$$E(R_v) = H_v \le \log v + 1,$$

we have, by Jensen's inequality and (2.1),

$$\begin{aligned} E(B_{n1}) &\le 1 + E(\log V_{n1}) \le 1 + \log(E(V_{n1})) \\ &\le 1 + \log(H_n) \qquad (2.5) \\ &= O(\log\log n). \end{aligned}$$

Furthermore, $(R_v - \log v)/(\log v)^{1/2} \Rightarrow \mathscr{N}(0,1)$ (see Durrett [5], page 102, for instance). A simple conditioning argument based on (2.4) and a trivial fact, that $V_n \le n$, leads then to $(B_n - \log n)/(\log n)^{1/2} \Rightarrow \mathscr{N}(0,1)$. It follows that $(t_n(w_1) - \log n)/(\log n)^{1/2} \Rightarrow \mathscr{N}(0,1)$ too, because by (2.5), in probability,

$$\lim \frac{B_{n1}}{(\log n)^{1/2}} = 0. \qquad \square$$

Consider $\Sigma_n = \Sigma_w t_n(w)$, which is the total number of all stable pairs (m, w).

COROLLARY 2.2. *$E(\Sigma_n)$ is asymptotic to $n \log n$.*

A simple proof is based on Theorem 2.1 and an inequality $E(t_n(w)) \le E(R_n)$. We conjecture that, in fact, $\Sigma_n / n \log n \to 1$, in probability.

3. Distribution of men's and women's ranks in stable matchings. By symmetry, each one of $n!$ matchings has the same probability P_n of being stable. Knuth [17] established an integral formula,

$$(3.1) \qquad P_n = \overbrace{\int \cdots \int_C}^{2n} \prod_{1 \le i \ne j \le n} (1 - x_i y_j)\, dx\, dy,$$

where $dx = dx_1 \cdots dx_n$, $dy = dy_1 \cdots dy_n$ and

$$C = \left\{ (x, y) \colon 0 \le x_i \le 1, 0 \le y_j \le 1, 1 \le i, j \le n \right\}.$$

The proof was based on an inclusion–exclusion formula and an ingenious interpretation of each term as the value of a $2n$-dimensional integral with the integrand equal to the corresponding term in the expansion of the product $\prod_{1 \le i \ne j \le n}(1 - x_i y_j)$.

Using a different method, we obtained in [22] a more general formula for $P_n(k)$, the probability that a fixed matching M is stable and $Q(M) = k$ [$R(M) = k$, respectively]. Here, $Q(M)$ and $R(M)$ are the total rank of wives and the total rank of husbands in M, respectively. It is possible to get a similar formula for $P_n(k, l)$, the probability that M is stable and $Q(M) = k$, $R(M) = l$.

LEMMA 3.1. *For* $n \le k, l \le n^2$,

$$P_n(k,l) = \int \underset{C}{\cdots} \int [\xi^{k-n}\eta^{l-n}] \prod_{i,j} \left(\bar{x}_i\bar{y}_j + x_i\bar{y}_j\xi + \bar{x}_i y_j \eta\right) dx\, dy, \tag{3.2}$$

where $\bar{x}_i = 1 - x_i$, $\bar{y}_j = 1 - y_j$ *and the integrand equals the coefficient of* $\xi^{k-n}\eta^{l-n}$ *in the product, which is taken over* $1 \le i \ne j \le n$.

NOTE. Summing up the integrands in (3.2) over l and then over k, we get, in particular,

$$P_n(k) = \int \underset{C}{\cdots} \int [\xi^{k-n}] \prod_{i,j} (\bar{x}_i + x_i\bar{y}_j\xi)\, dx\, dy \tag{3.3}$$

(cf. [22]) and, since $\bar{x}_i + x_i\bar{y}_j = 1 - x_i y_j$,

$$P_n = \int \underset{C}{\cdots} \int \prod_{i,j} (1 - x_i y_j)\, dx\, dy$$

[see (3.1)].

PROOF OF LEMMA 3.1. (a) Let $(m_1, \dots, m_n)$ and $(w_1, \dots, w_n)$ be the arbitrarily ordered sets of men and women, respectively. The random ranking system can be generated as follows. We introduce two $n \times n$ random matrices $X = [X_{ij}]$, $Y = [Y_{ij}]$, whose entries are independent and uniform on $[0, 1]$, and we postulate that man m_i (woman w_j) ranks women (men) in the *increasing* order of entries in the ith row of X (the jth column of Y). As a result, each person's ranking of members of the opposite sex is chosen according to the uniform distribution on the set of all $n!$ rankings (permutations), and the rankings by different persons are independent.

(b) Without loss of generality, we may consider a standard matching M_0 formed by n pairs (m_i, w_i). By the definition of the ranking procedure given previously, M_0 is stable iff there does not exist a pair (i, j), $i \ne j$, such that $X_{ij} < X_{ii}$ *and* $Y_{ij} < Y_{jj}$. The wives' and the husbands' ranks are given by

$$\begin{aligned} Q(M_0) &= n + \sum_{i=1}^{n} |\{j: X_{ij} < X_{ii}\}|, \\ R(M_0) &= n + \sum_{j=1}^{n} |\{i: Y_{ij} < Y_{jj}\}|. \end{aligned} \tag{3.4}$$

Since all $X_{\alpha\beta}$, $Y_{\alpha\beta}$ are independent, we need only show that $P_n(k, l|x, y)$, the conditional probability of the event $\{M_0$ is stable, $Q(M_0) = k$, $R(M_0) = l\}$, given $X_{ii} = x_i$, $Y_{jj} = y_j$, $1 \le i$, $j \le n$, equals the integrand in (3.2). To this end, introduce $\chi(M_0)$, the indicator of the event $\{M_0$ is stable$\}$, and write

$$P_n(k, l|x, y) = \left[\xi^k \eta^l\right] E\left(\chi(M_0) \xi^{Q(M_0)} \eta^{R(M_0)} | \cdot\right). \tag{3.5}$$

Here and in the following, $E(\ |\cdot)$ and $P(\ |\cdot)$ stand for the conditional expectation and the conditional probability, given $X_{ii} = x_i$, $Y_{jj} = y_j$. To evaluate the conditional expectation in (3.5), introduce the following "marking–coloring" procedure. Let $\xi, \eta \in [0, 1]$ be fixed. Sift through the pairs (i, j), $i \ne j$, in any order. Whenever $X_{ij} < X_{ii}$ $(= x_i)$, *mark* (i, j) with probability ξ; whenever $Y_{ij} < Y_{jj}$ $(= y_j)$, *color* (i, j) with probability η. Assume that the coloring–marking operations for different pairs are independent. Only to define the procedure completely, assume also that if $X_{ij} < X_{ii}$ *and* $Y_{ij} < Y_{jj}$, then marking and coloring of (i, j) are done independently. [No such pair (i, j) exists on the event $\{M_0$ is stable$\}$.] Then [see (3.4) and (3.5)]

$$E\left(\chi(M_0) \xi^{Q(M_0)} \eta^{R(M_0)} | \cdot\right) = \xi^n \eta^n P(B| \cdot),$$

where B is the event $\{M_0$ is stable, and all pairs (i, j) which "qualify" for marking or coloring are marked or colored$\}$. Clearly, $B = \bigcap_{i \ne j} B_{ij}$, where B_{ij} is the event

$$\Big\{(X_{ii} < X_{ij}, Y_{jj} < Y_{ij}) \text{ or } \left(X_{ii} > X_{ij}, Y_{jj} < Y_{ij} \text{ and } (i, j) \text{ is marked}\right) \\ \text{ or } \left(X_{ii} < X_{ij}, Y_{jj} > Y_{ij} \text{ and } (i, j) \text{ is colored}\right)\Big\}.$$

Notice now that, given $X_{\alpha\alpha} = x_\alpha$, $Y_{\beta\beta} = y_\beta$, $1 \le \alpha$, $\beta \le n$, the events B_{ij} are independent and

$$P\left(B_{ij}| \cdot\right) = \bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta.$$

Therefore, for ξ, $\eta \in [0, 1]$ and thus for all ξ, η,

$$E\left(\chi(M_0) \xi^{Q(M_0)} \eta^{R(M_0)} | \cdot\right) = \xi^n \eta^n \prod_{i \ne j} \left(\bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta\right). \tag{3.6}$$

So [see (3.5)], the relation (3.2) follows. □

Let (M, M') be a *given* pair of matchings such that M' is obtained from M by breaking up some pairs $(m_{i_1}, w_{i_1}) \cdots (m_{i_r}, w_{i_r})$ in M and pairing m_{i_s} with $w_{i_{s+1}}$ $(w_{i_{r+1}} = w_{i_1})$. Let A denote the event that M is stable, and (m_{i_s}, w_{i_s}), $1 \le s \le r$, form a rotation in M, so that M' is stable, too. By symmetry again, $P(A)$ depends only on n and r, so we denote it by P_{nr}. Introduce also $P_{nr}(k, l)$, the probability of $A \cap \{Q(M') = k$ and $R(M) = l\}$ (this probability depends only on n, r, k and l).

Lemma 3.2. (a)

$$P_{nr} = \int \underset{C}{\cdots} \int \prod_{i=1}^{r} (x_i y_i) \prod_{i,j} (1 - x_i y_j)\, dx\, dy, \tag{3.7}$$

and (b) *for* $n' \le k,\ l \le n^2$,

$$P_{nr}(k,l) = \int \cdots_C \int \prod_{i=1}^{r} (x_i y_i)\left[\xi^{k-n'}\eta^{l-n'}\right] \times \prod_{i,j} \left(\bar{x}_i\bar{y}_j + x_i\bar{y}_j\xi + \bar{x}_i y_j\eta\right) dx\,dy, \qquad n' = n + r. \tag{3.8}$$

In both (3.7) *and* (3.8), *the second product is taken over* $1 \le i \ne j \le n$, *and* $j \ne i + 1 \pmod r$ *for* $1 \le i \le r$.

PROOF. Let M be the standard matching M_0 and let $\{(m_{i_s}, w_{i_s})\}_{1 \le s \le r}$ be $\{(m_i, w_i)\}_{1 \le i \le r}$. On the event A, we have the following, by the definition of a rotation (see the Introduction): for $1 \le i,\ j \le r$,

$$X_{ii} < X_{i,i+1}, \qquad Y_{i,i+1} < Y_{i+1,i+1}, \tag{3.9}$$

$$i \ne j, \qquad X_{ij} < X_{i,i+1} \quad \Rightarrow \quad Y_{ij} > Y_{jj}. \tag{3.10}$$

Of course, for all $1 \le i,\ j \le n$,

$$X_{ij} < X_{ii} \quad \Rightarrow \quad Y_{ij} > Y_{jj}, \tag{3.11}$$

since on the event A the matching M_0 is stable. To see why (3.9) and (3.10) are valid, recall that on the event A the man m_i prefers his wife in M_0 to his new wife w_{i+1} in M_0', and w_{i+1} likes m_i more than her old husband m_{i+1}, thus the condition (3.9). As for the condition (3.10), it follows from an observation that (on A) w_{i+1} is the best woman for the man m_i among those women who prefer m_i to their husbands in M_0. Consequently, $Q(M_0')$ and $R(M_0)$ (the wives' rank in M_0' and the husbands' rank in M_0) are given by

$$Q(M_0') = n' + \sum_{i=1}^{r} \left|\{j \ne i: X_{ij} < X_{i,i+1}\}\right| + \sum_{i=r+1}^{n} \left|\{j: X_{ij} < X_{ii}\}\right|, \tag{3.12}$$

and

$$R(M_0) = n' + \sum_{j=1}^{r} \left|\{i \ne j-1: Y_{ij} < Y_{jj}\}\right| + \sum_{j=r+1}^{n} \left|\{i: Y_{ij} < Y_{jj}\}\right|, \tag{3.13}$$

where $n' = n + r$ and $j - 1 = r$ for $j = 1$.

Denote by $P_{nr}(k, l|\mathbf{x}, \mathbf{y})$ the conditional probability of the event $A \cap \{Q(M_0') = k,\ R(M_0) = l\}$, given

$$\begin{aligned} X_{i,i+1} &= x_i, \qquad X_{ii} = x_i', \qquad 1 \le i \le r, \\ X_{ii} &= x_i, \qquad r+1 \le i \le n, \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} Y_{j-1,j} &= y_j', \qquad 1 \le j \le r, \\ Y_{jj} &= y_j, \qquad 1 \le j \le n. \end{aligned} \tag{3.15}$$

Of course, $x_i' \le x_i$, $1 \le i \le r$, and $y_j' \le y_j$, $1 \le j \le r$ [see (3.9)].

Now (cf. the proof of Lemma 3.1),

$$P_{nr}(k, l|\mathbf{x}, \mathbf{y}) = [\xi^k \eta^l] E(\chi(A) \xi^{Q(M_0')} \eta^{R(M_0)} | \cdot) \tag{3.16}$$

and [see (3.12) and (3.13)]

$$E(\chi(A) \xi^{Q(M_0')} \eta^{R(M_0)} | \cdot) = \xi^{n'} \eta^{n'} P(\mathscr{B} | \cdot), \qquad \xi, \eta \in [0,1]. \tag{3.17}$$

Here $\mathscr{B}$ is the event $A \cap$ {all eligible pairs are marked or colored}, and we call a pair $(i, j) \notin M_0 \cup M_0'$ eligible for marking (coloring) if $X_{ij} < x_i$ and $(i, j) \neq (i, i)$ for $i \le r$ [if $Y_{ij} < y_j$ and $(i, j) \neq (j-1, j)$ for $j \le r$, respectively]. Marking (coloring) of an eligible pair (i, j) is done independently of all other such pairs including coloring (marking) of (i, j) itself, with probability ξ (probability η, respectively). Using the conditions (3.10) and (3.11), which mean, in particular, that on the event A no (i, j) is eligible for both marking and coloring, we get then

$$P(\mathscr{B} | \cdot) = \prod_{(i,j)} \left(\bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta \right). \tag{3.18}$$

[The product is over $(i, j) \notin M_0 \cup M_0'$.]

Plugging (3.17) and (3.18) into (3.16), we have

$$P_{nr}(k, l|\mathbf{x}, \mathbf{y}) = [\xi^{k-n'} \eta^{l-n'}] \prod_{(i,j)} \left(\bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta \right),$$

thus is *independent* of x', y'. It remains to notice that, fixing x, y and integrating $P_{nr}(k, l|\mathbf{x}, \mathbf{y})$ over x', y' subject to the restrictions $x_i' \le x_i$, $y_j' \le y_j$, $1 \le i, j \le r$, we obtain the integrand in (3.8).

Relation (3.7) follows from (3.8). □

NOTE. Contrary to the first impression, the integral formulas (3.2) (Lemma 3.1) and (3.7) and (3.8) (Lemma 3.2) are quite amenable to asymptotic study, primarily due to the product-type nature of the integrands.

4. Law of hyperbola. Let $\mathscr{M}$ denote a generic stable matching in the random instance of the stable marriage problem. This "law of hyperbola" states that, uniformly for all $\mathscr{M}$, the product $Q(\mathscr{M})R(\mathscr{M})$ is asymptotic to n^3, with probability approaching 1 exponentially fast. Here is the precise statement.

THEOREM 4.1. *Given $\delta \in (0,1)$, consider an event*

$$B_\delta = \{|n^{-3} Q(\mathscr{M}) R(\mathscr{M}) - 1| \ge \delta, \text{ for some } \mathscr{M}\}.$$

For all $c < c(\delta)$ and $n \ge n(\delta)$,

$$P(B_\delta) \le e^{-cn};$$

here

$$c(\delta) = \min(d(\delta), d(-\delta))$$

and

(4.1) $d(\theta) = \theta^{-1}\log(1+\theta) - 1 - \log(\theta^{-1}\log(1+\theta)), \qquad \theta > -1.$

Since $d(\theta) > 0$ for $\theta \neq 0$, $P(B_\delta)$ is exponentially small.

PROOF. Introduce the events

$$B_\delta^+ = \{n^{-3}Q(\mathscr{M})R(\mathscr{M}) \geq 1 + \delta, \text{ for some } \mathscr{M}\},$$
(4.2)
$$B_\delta^- = \{n^{-3}Q(\mathscr{M})R(\mathscr{M}) \leq 1 - \delta, \text{ for some } \mathscr{M}\}$$

and their probabilities $P_\delta^\pm$. Consider L, the set of pairs (k, l) of integers such that $n \leq k,\ l \leq n^2$. If $(k_1, l_1) \in L$ is such that $k_1 l_1 \geq (1+\delta)n^3$, then there exists $(k, l) \in L$ such that $k = k_1$, $l \leq l_1$ and

$$kl \geq (1+\delta)n^3, \qquad k(l-1) \leq (1+\delta)n^3,$$

that is,

$$(1+\delta)n^3 \leq kl \leq (1+\delta+n^{-1})n^3. \tag{4.3}$$

Similarly, if $(k_2, l_2) \in L$ satisfies $k_2 l_2 \leq (1-\delta)n^3$, then there exists $(k, l) \in L$ such that $k = k_2$, $l \geq l_2$ and

$$(1-\delta-n^{-1})n^3 \leq kl \leq (1-\delta)n^3. \tag{4.4}$$

Consider L_δ^+ (L_δ^-), the set of all (k, l) from L which satisfy condition (4.3) [condition (4.4), respectively]. For $(k, l) \in L_\delta^+$ (L_δ^-) define $P_n^+(k, l)$ $[P_n^-(k, l)]$ as the probability that the standard matching M_0 is stable and $Q(M_0) \geq k$, $R(M_0) \geq l$ $[Q(M_0) \leq k,\ R(M_0) \leq l$, respectively]. Then, appealing to the preceding argument, we have

$$P_\delta^+ \leq n! \sum_{L_\delta^+} P_n^+(k, l), \tag{4.5}$$

$$P_\delta^- \leq n! \sum_{L_\delta^-} P_n^-(k, l). \tag{4.6}$$

[Indeed, (4.5) states that P_δ^+ is at most the expected number of stable matchings $\mathscr{M}$ such that $Q(\mathscr{M}) \geq k$, $R(\mathscr{M}) \geq l$, $(k, l) \in L_\delta^+$.]

In the spirit of Chernoff's method [2], we can estimate the probabilities $P_n^\pm(k, l)$ by using Lemma 3.1:

$$P_n^+(k, l) \leq \int_C \inf\{\Phi(\xi, \eta, x, y) : \xi, \eta \geq 1\}\, dx\, dy, \tag{4.7}$$

$$P_n^-(k, l) \leq \int_C \inf\{\Phi(\xi, \eta, x, y) : \xi, \eta \in (0, 1]\}\, dx\, dy, \tag{4.8}$$

where

$$\Phi(\xi, \eta, x, y) = \xi^{n-k}\eta^{n-l} \prod_{i \neq j} \left(\bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta\right). \tag{4.9}$$

Working with these integrals, we shall assume (without loss of generality) that $k \le l$, so that in both (4.7) and (4.8)

$$k \le c_1 n^{3/2}, \qquad l \ge c_2 n^{3/2}, \qquad c_i = c_i(\delta) > 0. \tag{4.10}$$

(a) Fix $\gamma \in (\frac{1}{2}, 1)$, $\rho \in (0,1)$. Introduce

$$C_1 = \left\{ (x,y) \in C \colon \sum_{i=1}^{n} x_i y_i \ge n^{\gamma} \right\},$$

$$C_2 = \left\{ (x,y) \in C \colon \sum_{i=1}^{n} x_i \ge n - n^{\rho} \right\},$$

$$C_0 = C \setminus (C_1 \cup C_2),$$

and denote by $I^{\pm}(C_\alpha)$ the contributions of C_α to the value of the integrals in (4.7) and (4.8) ($\alpha = 0, 1, 2$). Choosing $\xi = \eta \equiv 1$,

$$I^{\pm}(C_\alpha) \le \int_{C_\alpha} \prod_{i \ne j} (1 - x_i y_j)\, dx\, dy, \qquad \alpha = 1, 2,$$

where $(1 - \lambda \le e^{-\lambda})$,

$$\begin{aligned} \log \prod_{i \ne j} (1 - x_i y_j) &\le - \sum_{i \ne j} x_i y_j \\ &\le n - \Big(\sum_i x_i \Big) \Big(\sum_j y_j \Big). \end{aligned} \tag{4.11}$$

For C_1, by the Cauchy–Schwarz inequality,

$$\begin{aligned} \Big(\sum_i x_i \Big) \Big(\sum_j y_j \Big) &\ge \Big(\sum_i x_i^{1/2} y_i^{1/2} \Big)^2 \\ &\ge \Big(\sum_i x_i y_i \Big)^2 \ge n^{2\gamma}. \end{aligned}$$

So,

$$I^{\pm}(C_1) \le \exp(n - n^{2\gamma}), \tag{4.12}$$

and the total contribution of C_1 to $n! \sum_{L_\delta^{\pm}} P_n^{\pm}(k,l)$ is *superexponentially* small, since $2\gamma > 1$.

For C_2, using (4.11) and integrating innermost over y_j, $1 \le j \le n$, we have

$$\begin{aligned} I^{\pm}(C_2) &\le e^n \int_{\{x:\, s \ge s_*\}} \left(\int_0^1 e^{-s y_1}\, dy_1 \right)^n dx \\ &= e^n \int_{s_*}^{n} \left(\frac{1 - e^{-s}}{s} \right)^n f_n(s)\, ds. \end{aligned} \tag{4.13}$$

Here $s = \sum_i x_i$, $s_* = n - n^{\rho}$ and $f_n(s)$ is the density of $\sum_{i=1}^{n} X_{ii}$, the sum of n independent, uniformly $[0,1]$ distributed random variables. It is well known [9] that

$$f_n(s) = \frac{s^{n-1}}{(n-1)!} P\Big(\max_{1\le i\le n} \mathscr{L}_i \le s^{-1}\Big), \tag{4.14}$$

where $\mathscr{L}_1, \dots, \mathscr{L}_n$ are the lengths of the consecutive subintervals of $[0,1]$ obtained by selecting independently $n-1$ points, each uniformly distributed on $[0,1]$. We show in the Appendix that

$$P\Big(\max_{1\le i\le n} \mathscr{L}_i \le \zeta\Big) \le (n\zeta - 1)^{n-1}, \qquad \forall\, \zeta \ge n^{-1}. \tag{4.15}$$

It follows then from (4.13)–(4.15) that

$$\begin{aligned} I^{\pm}(C_2) &\le \frac{e^n}{(n-1)!}\left(\frac{n^\rho}{n-n^\rho}\right)^{n-1} \int_{n-n^\rho}^{n} s^{-1}\, ds \\ &= O\big((n!)^{-1} e^{(1+c)n} n^{(\rho-1)n}\big), \qquad \forall\, c > 0. \end{aligned}$$

Therefore, the *total* contribution of C_2 to $n!\Sigma_{L_\delta^{\pm}} P_n^{\pm}(k,l)$ is also *superexponentially* small, since $\rho < 1$.

(b) Let $(x,y) \in C_0$. Similarly to (4.11), we obtain

$$\begin{aligned} &\log \prod_{i\ne j} \big(\bar{x}_i\bar{y}_j + x_i\bar{y}_j\xi + \bar{x}_i y_j \eta\big) \\ &\quad\le \sum_{i\ne j} \big(x_i(\xi-1) + y_j(\eta-1) + x_i y_j(1-\xi-\eta)\big) \\ &\quad= (n-1)(\xi-1)s - s(\xi,\eta)\sum_j y_j + (\xi+\eta-1)\sum_i x_i y_i, \end{aligned} \tag{4.16}$$

where

$$s(\xi,\eta) = s\xi - (n-1-s)(\eta-1). \tag{4.17}$$

By the definition of C_0, the last term in (4.16) is of order $O(|\xi+\eta-1| \times (n^\gamma \wedge s))$, where $n^\gamma \wedge s = \min(n^\gamma, s)$. So, $\Phi(\cdot)$ in (4.9) is estimated via

$$\begin{aligned} \Phi(\xi,\eta,x,y) \le \xi^{n-k}\eta^{n-l} \exp\Big[&(n-1)(\xi-1)s - s(\xi,\eta)\sum_j y_j \\ &+ O(|\xi+\eta-1|(n^\gamma \wedge s))\Big]. \end{aligned} \tag{4.18}$$

To estimate $I^{\pm}(C_0)$, we choose $\xi = \xi^{\pm}(s)$, $\eta = \eta^{\pm}(s)$, postponing the exact definitions until later. Clearly,

$$C_0 \subseteq \big\{(x,y)\colon s \le s_*,\ 0 \le y_j \le 1,\ 1 \le j \le n\big\};$$

so, integrating innermost with respect to y, we obtain

$$I^{\pm}(C_0) \le J^{\pm}(C_0),$$

$$\begin{aligned} J^{\pm}(C_0) = ((n-1)!)^{-1} \int_0^{s_*} \exp\big(&H(\xi^{\pm},\eta^{\pm},s) \\ &+ O(|\xi^{\pm} + \eta^{\pm} - 1|(n^\gamma \wedge s))\big)\, ds. \end{aligned} \tag{4.19}$$

Here

$$H(\xi,\eta,s) = (n-k)\log\xi + (n-l)\log\eta + (n-1)(\xi-1)s \tag{4.20}$$
$$+ n\log\frac{1-\exp(-s(\xi,\eta))}{s(\xi,\eta)} + (n-1)\log s,$$

$$s(\xi,\eta) = s\xi - (n-1-s)(\eta-1), \tag{4.21}$$

and we have used an inequality,

$$f_n(s) \le \frac{s^{n-1}}{(n-1)!}$$

[see (4.14)]. Whatever our choice of functions $\xi^{\pm}(s), \eta^{\pm}(s)$ is going to be, they must satisfy the basic restrictions $\xi^{+}(s), \eta^{+}(s) \ge 1$, $s \in (0, s_*)$, and $\xi^{-}(s)$, $\eta^{-}(s) \in (0,1]$, $s \in (0, s_*)$.

To describe the functions $\xi^{\pm}, \eta^{\pm}$ which we use in (4.19), we need to define three numbers $t_1, t_2, t_3 \in (0, s_*)$; t_1 and t_2 are given by

$$t_1 = \frac{k}{n}, \qquad t_2 = \frac{n^2}{l}.$$

By (4.10), $t_1, t_2 = O(n^{1/2})$, and also $t_1/t_2 = kl/n^3$. So,

$$t_1/t_2 = 1 + \delta_n, \qquad \delta_n = \begin{cases} \delta + O(n^{-1}), & \text{for } J^+, \\ -\delta + O(n^{-1}), & \text{for } J^-. \end{cases} \tag{4.22}$$

In particular, $t_1 > t_2$ for J^+, and $t_1 < t_2$ for J^-. Define t_3 by

$$t_3 = \frac{\log(1+\delta_n)}{\delta_n} t_1; \tag{4.23}$$

it is easy to check that for both J^+ and J^-, t_3 lies between t_1 and t_2.

Now, here are the functions $\xi^{\pm}, \eta^{\pm}$:

$$\xi^{+}(s) = \begin{cases} 1, & s \ge t_3, \\ \dfrac{t_1}{s}, & s < t_3, \end{cases} \qquad \eta^{+}(s) = \begin{cases} \dfrac{n-1-t}{n-1-s}, & s \ge t_3, \\ 1, & s < t_3, \end{cases} \tag{4.24}$$

$$\xi^{-}(s) = \begin{cases} \dfrac{t_1}{s}, & s \ge t_3, \\ 1, & s < t_3, \end{cases} \qquad \eta^{-}(s) = \begin{cases} 1, & s \ge t_3, \\ \dfrac{n-1-t_2}{n-1-s}, & s < t_3. \end{cases} \tag{4.25}$$

Obviously, the basic restrictions are met. The presence of functions $\xi(s) = t_1/s$, $\eta(s) = (n-1-t_2)/(n-1-s)$ can be explained right away by an observation that they happen to be the asymptotic solutions of the equations

$$H_\xi(\xi,1,s) = 0, \qquad H_\eta(1,\eta,s) = 0,$$

in the case $t_1 \to \infty$. The choice of the break-up point t_3 will become clear when we start estimating $H(\xi^{\pm}, \eta^{\pm}, s)$.

Let us consider the case of J^+. Suppose that $s \in [t_3, s_*]$. Then [see (4.17), (4.18) and (4.24)] $s(\xi, \eta) = t_2$ and

$$H(\xi^+, \eta^+, s) = (n - l)\log \frac{n - 1 - t_2}{n - 1 - s} + n \log \frac{1 - e^{-t_2}}{t_2} + (n - 1)\log s.$$

Here $n - l < 0$, $t_2 = n^2/l = O(n^{1/2})$; so, using

$$\log(n - 1 - t_2) = \log(n - 1) - \frac{t_2}{n - 1} + O\left(\left(\frac{t_2}{n - 1}\right)^2\right),$$

$$\log \frac{1}{n - 1 - s} \geq -\log(n - 1) + \frac{s}{n - 1},$$

we obtain

$$H(\xi^+, \eta^+, s) \leq nf(s) + O(n^{1/2}),$$

where

$$f(s) = \frac{t_2 - s}{t_2} + \log \frac{s}{t_2}.$$

The function f is concave down, and it achieves its absolute maximum (which is zero) at $s = t_2$. Therefore,

$$\max\{H(\xi^+, \eta^+, s): s \in [t_3, s_*]\} \leq nf(t_3) + O(n^{1/2}). \tag{4.26}$$

Suppose that $s \in (0, t_3]$. Then [again by (4.17), (4.18) and (4.24)] $s(\xi, \eta) = t_1$ and

$$\begin{aligned} H(\xi^+, \eta^+, s) &= (n - k)\log \frac{t_1}{s} + (n - 1)(t_1 - s) \\ &\quad + n \log \frac{1 - e^{-t_1}}{t_1} + (n - 1)\log s \\ &\leq (k - 1)g(s), \end{aligned}$$

where

$$g(s) = \frac{t_1 - s}{t_1} + \log \frac{s}{t_1}.$$

Like f, the function g is concave down, and it achieves its zero maximum at $s = t_1$. Since $k \geq (1 + \delta)n^3/l \geq (1 + \delta)n$, we have

$$\max\{H(\xi^+, \eta^+, s): s \in (0, t_3]\} \leq ng(t_3).$$

Now it is time to explain our choice of t_3 in (4.22). It satisfies $g(s) = f(s)$. Simple computations show that the common value of $g(t_3)$ and $f(t_3)$ is

$$-d(\delta_n) = -d(\delta) + O(n^{-1}) \tag{4.27}$$

[see (4.1) for the definition of the function $d(\cdot)$]. Thus

$$\max\{H(\xi^+, \eta^+, s) : s \in (0, s_*)\} \le -nd(\delta) + O(n^{1/2}). \tag{4.28}$$

Let us estimate the remainder term of the integrand in (4.19). If $t_3 \le s \le s_*$ $(= n - n^\rho)$, then [see (4.24)]

$$|\xi^+ + \eta^+ - 1|(n^\gamma \wedge s) \le \frac{n - 1 - t_2}{n - 1 - s} n^\gamma = O(n^\sigma); \tag{4.29}$$

here $\sigma = 1 + \gamma - \rho$, and σ can be made arbitrarily close to $\frac{1}{2}$ (from above) by choosing γ $(> \frac{1}{2})$ and ρ (< 1) sufficiently close to their respective bounds. If $0 < s \le t_3$, then [see (4.24)]

$$|\xi^+ + \eta^+ - 1|(n^\gamma \wedge s) \le \xi^+(s)s = t_1 = O(n^{1/2}). \tag{4.30}$$

Combining (4.19) and (4.28)–(4.30) enables us to conclude that

$$J^+(C_0) = O\big(((n-1)!)^{-1} s_* \exp(-nd(\delta) + O(n^\sigma))\big).$$

Thus, the total contribution of C_0 to the value of $n!\sum_{L_\delta^+} P_n^+(k, l)$ is bounded by

$$O(ns_* \exp(-nd(\delta) + O(n^\sigma))), \qquad \sigma = 1 + \gamma - \rho.$$

Since the contributions of C_1, C_2 are superexponentially small for every $\gamma \in (\frac{1}{2}, 1)$ and $\rho \in (0, 1)$, we see that

$$P_\delta^+ \le n! \sum_{L_\delta^+} P_n^+(k, l) = O(\exp(-nd(\delta) + O(n^\sigma))), \tag{4.31}$$

for arbitrary $\sigma \in (\frac{1}{2}, 1)$.

In exactly the same way, it could be shown that

$$P_\delta^- = O(\exp(-nd(-\delta) + O(n^\sigma))). \tag{4.32}$$

The estimates (4.31) and (4.32) imply the assertion of Theorem 4.1. □

NOTE. Suppose that $\delta = n^{-\lambda}$, $\lambda \in (0, \frac{1}{4})$. Since $\lambda < 1$, the preceding argument works without any changes. (We still have that $\delta_n > 0$, $t_2 < t_3 < t_1$ for J^+, and $\delta_n < 0$, $t_1 < t_2 < t_3$ for J^- [see (4.22) and (4.23)].) Since $d(\theta) \sim -\theta^2/8$ as $\theta \to 0$, instead of (4.27) we have

$$-d(\delta_n) = -d(n^{-\lambda} + O(n^{-1})) \sim -n^{-2\lambda}/8$$

and, similarly, for the "minus" case

$$-d(\delta_n) = -d(-n^{-\lambda} + O(n^{-1})) \sim -n^{-2\lambda}/8.$$

Thus, we have [cf. (4.31) and (4.32)] the bounds

$$P_\delta^\pm = O\big(\exp(-\tfrac{1}{8}n^{1-2\lambda} + O(n^\sigma))\big), \qquad \forall\, \sigma \in (0, \tfrac{1}{2}),$$

which are superpolynomially small since $1 - 2\lambda > \frac{1}{2}$. Hence, we have the following theorem.

THEOREM 4.1′. *For every* $\lambda \in (0, \frac{1}{4})$,

$$P(|n^{-3}Q(\mathscr{M})R(\mathscr{M}) - 1| \geq n^{-\lambda} \text{ for some } \mathscr{M}) = O(\exp(-\tfrac{1}{9}n^{1-2\lambda})).$$

NOTE. As Theorems 4.1 and 4.1′ indicate, the law of hyperbola holds asymptotically for almost all $n \times n$ instances of the problem. However, it is not difficult to construct special instances for which $Q(\mathscr{M})R(\mathscr{M})$ is as low as n^2 and as high as $n^2(n+1)^2/4$.

5. Likely rotations. Let $\mathscr{M}$, $\mathscr{M}'$ denote a pair of generic stable matchings such that $\mathscr{M}'$ is obtained from $\mathscr{M}$ by elimination of a rotation in $\mathscr{M}$. Let $r(\mathscr{M}, \mathscr{M}')$ stand for the number of pairs in $\mathscr{M}$ which form this rotation. Our next statement shows that with high probability for every such pair of stable matchings the length of the rotation is of order at most $(n \log n)^{1/2}$.

THEOREM 5.1. *For* $2 \leq r \leq n$, *denote by* π_{nr} *the probability that for some* $\mathscr{M}$, $\mathscr{M}'$ *the number* $r(\mathscr{M}, \mathscr{M}')$ *equals* r. *Then, uniformly over* r,

$$\pi_{nr} = O\left(\frac{n \log n}{r} \exp\left(-\frac{r^2}{n}\right)\right).$$

Consequently, for every $a > 0$,

$$P\left(\max_{(\mathscr{M}, \mathscr{M}')} r(\mathscr{M}, \mathscr{M}') \leq ((1+a)n \log n)^{1/2}\right) \geq 1 - O(n^{-a}). \tag{5.1}$$

PROOF. The probability π_{nr} is bounded by the expected number of the pairs $(\mathscr{M}, \mathscr{M}')$ such that $r(\mathscr{M}, \mathscr{M}') = r$. Since there are $\binom{n}{r}(r-1)!$ ways to select a cyclically ordered sequence of r pairs in a given matching, we have

$$\pi_{nr} \leq n!\binom{n}{r}(r-1)!P_{nr}. \tag{5.2}$$

Here P_{nr} is the probability that M_0 (the standard matching) is stable and that the pairs (m_i, w_i), $1 \leq i \leq r$, form a rotation in this matching. By Lemma 3.2(a),

$$P_{nr} = \int_C \prod_{k=1}^{r} (x_k y_k) \prod_{i,j} (1 - x_i y_j)\, dx\, dy; \tag{5.3}$$

in the second product, $i \neq j$ and $j \neq i + 1 \pmod r$ if $1 \leq i \leq r$. Introduce s_j, $1 \leq j \leq n$, as follows: $s_j = \sum_{i \neq j} x_i$ for $r + 1 \leq j \leq n$, and $s_j = \sum_{i \neq j, j-1} x_i$ for $1 \leq j \leq r$, $x_0 = x_r$. Then, bounding each factor $(1 - x_i y_j)$ by $\exp(-x_i y_j)$, rearranging factors and integrating with respect to $y_j \in [0, 1]$, $1 \leq j \leq n$, we can write

$$P_{nr} \leq \int_x \prod_{j=1}^{r} \Psi_2(s_j) \prod_{k=r+1}^{n} \Psi_1(s_k) \prod_{l=1}^{r} x_l\, dx,$$

where

$$\Psi_2(u) = u^{-2}(1 - e^{-u}(1 + u)), \tag{5.4}$$

$$\Psi_1(u) = u^{-1}(1 - e^{-u}). \tag{5.5}$$

A simple calculation shows that, for $u > 0$ and some constant $c > 0$,

$$(\log \Psi_1(u))' \leq -\frac{c}{u+1}, \qquad (\log \Psi_2(u))' \leq -\frac{c}{u+1}. \tag{5.6}$$

So, introducing $s = \sum_{i=1}^{n} x_i$, by the definition of s_j and the fact that each $x_j \in [0, 1]$,

$$\begin{aligned} \prod_{j=1}^{r} \Psi_2(s_j) \prod_{k=r+1}^{n} \Psi_1(s_k) &\leq \Psi_2(s)^r \Psi_1(s)^{n-r} \exp\left[\frac{c}{s}\left(\sum_{j=1}^{n} (x_j + x_{j-1})\right)\right] \\ &\leq c' \Psi_2(s)^r \Psi_1(s)^{n-r}, \end{aligned} \tag{5.7}$$

$c' = e^{2c}$. Thus,

$$P_{nr} \leq c' \int_{x \geq 0} \Psi_2(s)^r \Psi_1(s)^{n-r} \left(\prod_{l=1}^{r} x_l\right) dx. \tag{5.8}$$

To simplify this bound further, we need the relations

$$\int_{\{\sum_{l=1}^{r} x_l \leq t\}} \prod_{l=1}^{r} x_l \, dx_l = \frac{t^{2r}}{(2r)!}, \qquad x_1, \ldots, x_r \geq 0,$$

$$\int_{\{\sum_{k=r+1}^{n} x_k \leq u\}} \prod_{k=r+1}^{n} dx_k = \frac{u^{n-r}}{(n\text{-}r)!}, \qquad x_{r+1}, \ldots, x_n \geq 0.$$

Then (5.8) implies that

$$\begin{aligned} P_{nr} &\leq \frac{c'}{(2r-1)!(n-r-1)!} \\ &\quad \times \int_{t, u \geq 0} \Psi_2(s)^r \Psi_1(s)^{n-r} t^{2r-1} u^{n-r-1} \, dt \, du, \end{aligned} \tag{5.9}$$

where $s = t + u$. Now, for nonnegative integers ν, μ,

$$\int_{t+u \leq a} t^{\nu} u^{\mu} \, dt \, du = \frac{\nu! \mu!}{(\nu + \mu + 2)!} a^{\nu + \mu + 2}, \qquad t, u \geq 0.$$

So, transforming the integral in (5.9) into a one-dimensional integral, we obtain

$$P_{nr} \leq \frac{c'}{(n+r-1)!} \int_0^n \Psi_2(s)^r \Psi_1(s)^{n-r} s^{n+r-1} \, ds.$$

Thus, using (5.4) and (5.5) and throwing away $(1 - e^{-s})^{r-1}(1 - e^{-s}(1 + s))^{n-r}$, we have

$$P_{nr} \leq \frac{c'}{(n+r-1)!} \int_0^n \frac{1-e^{-s}}{s}\, ds = O\left(\frac{\log n}{(n+r-1)!}\right),$$

that is [by (5.2)],

$$\pi_{nr} = O\left(\frac{n!\binom{n}{r}(r-1)!}{(n+r-1)!} \log n\right). \tag{5.10}$$

Let us have a closer look at the last bound. Clearly,

$$\frac{n!\binom{n}{r}(r-1)!}{(n+r-1)!} = \frac{n}{r} \prod_{j=1}^{r-1} \frac{n-j}{n+j} = \frac{n}{r} \exp\left(\sum_{j=1}^{r-1} \log \frac{n-j}{n+j}\right).$$

Bounding the last sum by the integral leads, after some rearranging, to

$$\begin{aligned}\sum_{j=1}^{r-1} \log \frac{n-j}{n+j} &\leq \int_0^{r-1} \log \frac{n-z}{n+z}\, dz \\ &= -n\left[\left(1+\frac{r-1}{n}\right)\log\left(1+\frac{r-1}{n}\right)\right. \\ &\qquad \left. + \left(1-\frac{r-1}{n}\right)\log\left(1-\frac{r-1}{n}\right)\right].\end{aligned} \tag{5.11}$$

Since

$$(1+x)\log(1+x) + (1-x)\log(1-x) \geq x^2,$$

the bounds (5.10) and (5.11) imply that

$$\pi_{nr} = O\left(\frac{n \log n}{r} \exp\left(-\frac{r^2}{n}\right)\right). \tag{5.12}$$

The rest of the argument is short. Denote $r_0 = [((1+a)n \log n)^{1/2}]$. By (5.12), we can write

$$\begin{aligned} P\left(\max_{(\mathcal{M},\mathcal{M}')} r(\mathcal{M},\mathcal{M}') > r_0\right) &\leq \sum_{r>r_0} \pi_{nr} \\ &= O\left(n \log n \int_{r_0}^{\infty} z^{-1} \exp\left(-\frac{z^2}{n}\right) dz\right) \\ &= O\left(n \log n \frac{\exp(-r_0^2/n)}{r_0^2/n}\right) = O(n^{-a}). \end{aligned}$$

This completes the proof of Theorem 5.1. □

Since the likely rotations have lengths of order $O((n \log n)^{1/2}) = o(n)$, it seems plausible that, typically, the points $(Q(\mathscr{M}), R(\mathscr{M}))$ and $(Q(\mathscr{M}'), R(\mathscr{M}'))$ are not too far apart. As our next statement shows, this is, indeed, the case.

THEOREM 5.2. *Let* $\delta = n^{-\lambda}$, $\lambda \in (0, \frac{1}{4})$. *Consider an event* ${}_1B_\delta = \{|n^{-3}Q(\mathscr{M}')R(\mathscr{M}) - 1| \ge \delta$, *for some pair* $(\mathscr{M}, \mathscr{M}')$ *of one-rotation-connected stable matchings*$\}$. *Then*

$$P({}_1B_\delta) = O(n^{-a}),$$

for every $a > 0$. (*In terminology adopted in Knuth, Motwani and Pittel* [19], ${}_1\overline{B}_\delta$ *is a quite sure event.*)

PROOF. By Theorem 5.1, we need to prove only that

$$P\big({}_1B_\delta^{\pm} \cap A\big) = O(n^{-a}),$$

where A is the event

$$\Big\{\max_{(\mathscr{M}, \mathscr{M}')} r(\mathscr{M}, \mathscr{M}') \le r_0\Big\}, \qquad r_0 = \Big[\big((1+a)n \log n\big)^{1/2}\Big],$$

and ${}_1B_\delta^{\pm}$ are defined similarly to $B_\delta^{\pm}$ in (4.2), except that $Q(\mathscr{M})$ is replaced by $Q(\mathscr{M}')$. Like (4.5) and (4.6), we can write then

$$P\big({}_1B_\delta^{\pm} \cap A\big) \le n! \sum_{r \le r_0} \binom{n}{r} (r-1)! \sum_{L_\delta^{\pm}} P_{nr}^{\pm}(k, l). \tag{5.13}$$

Here $P_{nr}^{\pm}(k, l)$ is the probability of the event that the matching M_0 is stable, the first r pairs in M_0 form a rotation and, in the pair $(\mathscr{M}, \mathscr{M}') = (M_0, M_0')$, $Q(\mathscr{M}') \ge k$, $R(\mathscr{M}) \ge l$ for the plus case, and $Q(\mathscr{M}') \le k$, $R(\mathscr{M}) \le l$ for the minus case.

As for the probabilities $P_{nr}^{\pm}(k, l)$, we estimate them by using Lemma 3.2(b), dropping from the integrand in (3.8) the factor $\prod_{i=1}^{r} x_i y_i$. (This factor, which is at most 1, has already served us well in the proof of Theorem 5.1, and we do not need it any longer, now that we consider only the "small" values of r.) The inequalities are remarkably similar to (4.7) and (4.8), except that $\Phi(\xi, \eta, x, y)$ has been replaced now by $\Phi_1(\xi, \eta, x, y)$,

$$\Phi_1(\xi, \eta, x, y) = \xi^{n'-k} \eta^{n'-l} \prod_{i,j} \big(\bar{x}_i \bar{y}_j + x_i \bar{y}_j \xi + \bar{x}_i y_j \eta\big)$$

$[i \ne j$, and $j \ne i + 1 \pmod r$ for $1 \le i \le r]$, where $n' = n + r$. Introduce

$${}_1C_1 = \Big\{(x, y) \in C\colon \sum_{i=1}^{n} x_i y_i + \sum_{i=1}^{r} x_i y_{i+1} \ge n^{\gamma}\Big\},$$

$${}_1C_2 = \Big\{(x, y) \in C\colon \sum_{i=1}^{n} x_i \ge n - n^{\rho}\Big\},$$

$${}_1C_0 = C \setminus ({}_1C_1 \cup {}_1C_2).$$

Repeating almost verbatim part (a) in the proof of Theorem 4.1, we can show that the contributions of the domains ${}_1C_1$ and ${}_1C_2$ to the bounds of $n!\Sigma_{L_\delta^\pm} P_{nr}^\pm(k,l)$ are superexponentially small. Their overall contributions to the bounds for $P({}_1B_\delta^\pm \cap A)$ in (5.12) are also then superexponentially small, since

$$\sum_{r \le r_0} \binom{n}{r}(r-1)! = O(r_0 n^{r_0}) \tag{5.14}$$
$$= O\Big(r_0 \exp\big[(1+a)^{1/2} n^{1/2} (\log n)^{3/2}\big]\Big).$$

For the domain ${}_1C_0$, analogously to (4.18), we show easily that

$$\Phi_1(\xi,\eta,x,y) \le \xi^{n'-k}\eta^{n'-l}\exp\Bigg[(n_1-1)(\xi-1)s - s_1(\xi,\eta)\sum_j y_j + O(|\xi+\eta-1|(n^\gamma \wedge s))\Bigg],$$

$$s_1(\xi,\eta) = s\xi - (n_1-1-s)(\eta-1).$$

Here $n_1 = n$ when $\xi,\eta \ge 1$, and $n_1 = n-1$ when $\xi,\eta \le 1$. So, with some minor notational changes, we prove that, as in the case of $P(B_\delta^\pm)$ (see the proofs of Theorems 4.1 and 4.1′), the overall contributions of ${}_1C_0$ to the bounds of $n!\Sigma_{L_\delta^\pm} P_{nr}^\pm(k,l)$ are of order $O(\exp(-nd(\pm\delta + O(n^{-1})) + O(n^\sigma)))$, $\forall\, \sigma \in (\frac{1}{2},1)$. Since the last bound is uniform for $r \le r_0$, it implies [in conjunction with (5.14)] that

$$P({}_1B_\delta^\pm \cap A) = O\big(\exp(-n^{1-2\lambda}/8 + O(n^\sigma))\big), \qquad \forall\, \sigma \in \left(\tfrac{1}{2},1\right).$$

The proof of Theorem 5.2 is complete. □

Now it is time to recall that whp $(Q(\mathscr{M}_1), R(\mathscr{M}_1))$ and $(Q(\mathscr{M}_2), R(\mathscr{M}_2))$ are asymptotic to $(n\log n, n^2/\log n)$ and $(n^2/\log n, n\log n)$, respectively. Here $\mathscr{M}_1$, $\mathscr{M}_2$ are the male-optimal stable matching and the female-optimal stable matching (Pittel [22]). It is not difficult to extract from the proofs in [22] that, in fact, $Q(\mathscr{M}_1)$, $R(\mathscr{M}_2)$ are quite surely (q.s.) of order $n \log n$; more precisely, for every $\varepsilon < 1$, $a > 0$ and $n \ge n(a,\varepsilon)$:

$$P((1-\varepsilon)n\log n \le Q(\mathscr{M}_1), R(\mathscr{M}_2) \le (1+a)n\log n) \ge 1 - n^{-a}. \tag{5.15}$$

Combination of this result with Theorem 4.1 yields that

$$P\left(R(\mathscr{M}_1), Q(\mathscr{M}_2) \ge \frac{n^2}{(1+a)\log n}\right) \ge 1 - O(n^{-a'}), \qquad \forall\, a' < a. \tag{5.16}$$

With a slight abuse of terminology, we describe the situation by saying that q.s. the points $\omega(\mathscr{M}_1) = (n^{-3/2}Q(\mathscr{M}_1), n^{-3/2}R(\mathscr{M}_1))$ and $\omega(\mathscr{M}_2) = (n^{-3/2}Q(\mathscr{M}_2), n^{-3/2}R(\mathscr{M}_2))$ are of order $(n^{-1/2}\log n, n^{1/2}/\log n)$ and $(n^{1/2}/\log n, n^{-1/2}\log n)$, respectively.

Let us recall now that, according to Irving and Leather [14], for every instance of the stable marriage problem, there exists a sequence of stable matchings $\mathscr{M}_1 = \mathscr{M}^{(1)}, \mathscr{M}^{(2)}, \dots, \mathscr{M}^{(k)} = \mathscr{M}_2$ such that each $\mathscr{M}^{(i)}$ is obtained from $\mathscr{M}^{(i-1)}$ by elimination of a rotation in $\mathscr{M}^{(i-1)}$. As a geometric consequence, the point $(Q(\mathscr{M}^{(i)}), R(\mathscr{M}^{(i)}))$ lies southeast from the point $(Q(\mathscr{M}^{(i-1)}, R(\mathscr{M}^{(i-1)}))$.

Let $\delta = n^{-\lambda'}$, $\lambda' \in (0, \frac{1}{4})$. In the (u,v)-plane, introduce a hyperbola-shaped strip $D = \{\omega = (u,v)\colon u, v > 0, |uv - 1| < \delta\}$. By Theorem 4.1′, we can assert that, whatever the chain $(\mathscr{M}^{(1)}, \dots, \mathscr{M}^{(k)})$ is, the related sequence $\{\omega^{(i)} = (n^{-3/2}Q(\mathscr{M}^{(i)}), n^{-3/2}R(\mathscr{M}^{(i)}))\}_{i=1}^{k}$ lies, q.s., in the strip D. Consider a hyperbolic arc $\mathscr{C}_n = \{\omega = (u,v)\colon uv = 1,\ n^{-\lambda} \le u, v \le n^{\lambda}\}$, where $\lambda \in (0, \lambda')$. For every $\omega \in \mathscr{C}_n$, define an open neighborhood $N(\omega)$ of ω,

$$N(\omega) = \{\omega' \in D\colon u' < (1+\delta)u, v' < (1+\delta)v\}.$$

Since $\delta = n^{-\lambda'}$ and $\lambda' > \lambda$,

$$\sup\{d(\omega, \omega')\colon \omega' \in N(\omega)\} = O(n^{\lambda - \lambda'}) = o(1),$$

uniformly over $\omega \in \mathscr{C}_n$. [$d(\cdot, \cdot)$ is the distance (Euclidean distance, say) in the (u,v) plane.]

Suppose that $\omega^{(i)} \in D \setminus N(\omega)$ for all $1 \le i \le k$ and some $\omega \in \mathscr{C}_n$. Then there must exist i, $1 \le i \le k-1$, such that $\omega^{(i)}$ and $\omega^{(i+1)}$ are located northwest and southeast from $N(\omega)$, respectively. Consequently, $u^{(i+1)}v^{(i)} \ge (1+\delta)^2$, that is,

$$Q(\mathscr{M}^{(i+1)})R(\mathscr{M}^{(i)}) \ge n^3(1+\delta)^2.$$

According to Theorem 5.2, the chances of this happening are superpolynomially small. Thus, we have proved the following result.

THEOREM 5.3. *For $\omega \in \mathscr{C}_n$ denote by $d(\omega)$ the distance between ω and the set $\{\omega(\mathscr{M})\colon \mathscr{M}$ is stable$\}$. Then, for each $t < \frac{1}{4} - \lambda$,*

$$P(\max\{d(\omega)\colon \omega \in \mathscr{C}_n\} \le n^{-t}) \ge 1 - O(n^{-a}), \qquad \forall\, a > 0.$$

[Recall that $\mathscr{C}_n$ consists of points ω of the hyperbola $uv = 1$, with $n^{-\lambda} < u$, $v < n^{\lambda}$, $\lambda \in (0, \frac{1}{4})$.] Thus, quite surely, all the points of $\mathscr{C}_n$ are close to the set $\{\omega(\mathscr{M})\}$.

Now, $(1,1)$ is a point of $\mathscr{C}_n$ [for every $\lambda \in (0, \frac{1}{4})$]. So, Theorem 5.3 and Theorem 4.1′ taken together imply directly the following result.

COROLLARY 5.4. *Let $\mathscr{M}_3$ denote an egalitarian stable marriage, that is,*

$$Q(\mathscr{M}_3) + R(\mathscr{M}_3) = \min\{Q(\mathscr{M}) + R(\mathscr{M})\}.$$

Then, for every $t \in (0, \frac{1}{4})$,

$$P\left(\left|\frac{Q(\mathscr{M}_3) + R(\mathscr{M}_3)}{2n^{3/2}} - 1\right| \le n^{-t}\right) \ge 1 - O(n^{-a}), \qquad \forall\, a > 0.$$

[*In contrast, both* $Q(\mathscr{M}_1) + R(\mathscr{M}_1)$ *and* $Q(\mathscr{M}_2) + R(\mathscr{M}_2)$ *are asymptotic, quite surely, to* $n^2/\log n$.] *Moreover, for every* $a > 0$,

$$(5.17)\quad P\left(|n^{-3/2}Q(\mathscr{M}_3) - 1| + |n^{-3/2}R(\mathscr{M}_3) - 1| \le n^{-t}\right) \ge 1 - O(n^{-a})$$

as well.

6. Likely range of individual ranks in a stable matching. Let $\mathscr{M}$ be a stable matching. For each man m_i, $1 \le i \le n$, let $Q_i(\mathscr{M})$ denote the rank of m_i's wife, as ranked by the man himself. Similarly, for each woman w_j, $1 \le j \le n$, let $R_j(\mathscr{M})$ denote the rank of w_j's husband. Introduce $\mathscr{Q}(\mathscr{M}) = \max_i Q_i(\mathscr{M})$ and $\mathscr{R}(\mathscr{M}) = \max_j R_j(\mathscr{M})$, which are the largest rank of a wife and the largest rank of a husband in $\mathscr{M}$. Set $\mathscr{Q} = \min_{\mathscr{M}} \mathscr{Q}(\mathscr{M})$, $\mathscr{R} = \min_{\mathscr{M}} \mathscr{R}(\mathscr{M})$; by symmetry, $\mathscr{Q} \equiv_{\mathscr{D}} \mathscr{R}$. By the properties of the matchings $\mathscr{M}_1$, $\mathscr{M}_2$, we know that

$$\mathscr{Q} = \max_i Q_i(\mathscr{M}_1), \qquad \mathscr{R} = \max_j R_j(\mathscr{M}_2).$$

We can see also (Section 2) that $\mathscr{Q}$ is the largest number of proposals made by a man in the McVitie–Wilson algorithm, since $Q_i(\mathscr{M}_1)$ is the number of proposals by the man m_i. We know that $E[Q_i(\mathscr{M}_1)]$ is asymptotic to $\log n$. How large is $E(\mathscr{Q})$?

THEOREM 6.1. (a) *For every* $\varepsilon \in (0, 1)$,

$$(6.1)\quad P(\mathscr{Q} \ge (1-\varepsilon)\log^2 n) \ge 1 - \exp(-c\log^2 n), \qquad \forall\, c < c(\varepsilon),$$

where

$$(6.2)\quad c(\varepsilon) = \varepsilon + (1-\varepsilon)\log(1-\varepsilon).$$

(b) *For every* $a > 0$,

$$(6.3)\quad P(\mathscr{Q} \le (2+a)\log^2 n) \ge 1 - O(n^{-c}), \qquad \forall\, c < c(a),$$

where

$$(6.4)\quad c(a) = 2a\left[3 + (4a+9)^{1/2}\right]^{-1}.$$

Consequently, $E(\mathscr{Q})$ *is of order* $\log^2 n$, *as* $n \to \infty$.

NOTES. (i) We are tempted to conjecture that $\mathscr{Q}/\log^2 n$ converges, in probability, to a constant. If this is actually so, then the constant lies between 1 and 2. (*See the Note added in proof at the end of the text.*) (ii) Thus, the expected largest number of proposals made by a man is precisely of order $\log^2 n$. It is worth pointing out here that the expected largest number of proposals received by a woman is only of order $\log n$.

PROOF OF THEOREM 6.1. As $\mathscr{Q} \equiv_{\mathscr{D}} \mathscr{R}$, we may and shall consider bounds for $\mathscr{R}$.

Let us recall that the ranking of men by women and women by men is induced by two matrices, Y and X. In particular, if in a stable matching $\mathscr{M}$ the woman w_j is matched with man m_i, then $R_j(\mathscr{M})$ (the rank of m_i as ranked by w_j) equals the rank of Y_{ij} among $Y_{1j}, \dots, Y_{nj}$.

(a) The proof of the lower bound (6.1) consists of two steps. First, we prove that whp each stable matching has a pair (m_i, w_j) such that

$$Y_{ij} \geq (1 - \varepsilon_1)\frac{\log^2 n}{n}, \qquad \forall\, \varepsilon_1 \in (0,1). \tag{6.5}$$

Second, assuming that $\varepsilon_1 < \varepsilon$, we prove that whp *every* column of the matrix Y contains at least $(1 - \varepsilon)\log^2 n$ entries which are less than $(1 - \varepsilon_1)(\log^2 n)/n$. These two facts imply that whp $\mathscr{R} \geq (1 - \varepsilon)\log^2 n$.

STEP 1. Introduce $\pi_n(\varepsilon_1)$, the probability that for some stable matching $\mathscr{M} = \{(m_{i_k}, w_{j_k})\}_{k=1}^n$,

$$Y_{i_k j_k} < \alpha =_{\text{def}} (1 - \varepsilon_1)\frac{\log^2 n}{n}, \qquad 1 \leq k \leq n.$$

Then

$$\pi_n(\varepsilon_1) \leq n!\,P_n(\varepsilon_1), \tag{6.6}$$

where $P_n(\varepsilon_1)$ is the probability that the standard matching M_0 is stable, and $Y_{jj} < \alpha$, $1 \leq j \leq n$. Analogously to (3.1), we have

$$P_n(\varepsilon_1) = \int_{C(\varepsilon_1)} \prod_{i \neq j} (1 - x_i y_j)\, dx\, dy, \tag{6.7}$$

where

$$C(\varepsilon_1) = \big\{(x, y) : 0 \leq x_i \leq 1, 0 \leq y_j \leq \alpha, 1 \leq i, j \leq n\big\}.$$

Acting as in the derivation of (5.7) from (5.3), we transform (6.7) into

$$\begin{aligned} P_n(\varepsilon_1) &\leq c\int_x \left(\frac{1 - e^{-\alpha s}}{s}\right)^n dx \\ &= \frac{c}{(n-1)!}\int_0^n \frac{(1 - e^{-\alpha s})^n}{s} P\Big(\max_{1 \leq i \leq n} \mathscr{L}_i \leq s^{-1}\Big)\, ds, \end{aligned} \tag{6.8}$$

[cf. (4.13) and (4.14)]. For $\beta \in (1 - \varepsilon_1, 1)$, introduce

$$s_0 = \frac{\beta}{\alpha}\log n = \frac{\beta}{(1 - \varepsilon_1)}\frac{n}{\log n},$$

and break the integral (6.8) into two parts, $s \in [0, s_0]$ and $s \in [s_0, n]$. The first integral is bounded by

$$(1 - \exp(-\beta \log n))^{n-1} s_0 = O\big(\exp(-n^{1-\beta'})\big), \qquad \forall\, \beta' \in (\beta, 1). \tag{6.9}$$

In the second integral,

$$
\begin{aligned}
P\Big(\max_{1\le i\le n}\mathscr{L}_i \le s^{-1}\Big) &\le P\bigg(\max_{1\le i\le n}\mathscr{L}_i \le \frac{1-\varepsilon_1}{\beta}\frac{\log n}{n}\bigg)\\
&= O(\exp(-n^{\rho})), \qquad \forall\,\rho\in\bigg(0, 1-\frac{1-\varepsilon_1}{\beta}\bigg)
\end{aligned}
\tag{6.10}
$$

(see the Appendix). Thus, the integral from 0 to n is bounded by $\exp(-n^{\sigma})$, for each σ less than $\min(1-\beta, 1-(1-\varepsilon_1)/\beta)$. The last function of β achieves its maximum $1-(1-\varepsilon_1)^{1/2}$ at $\beta=(1-\varepsilon_1)^{1/2}$. Summarizing [see (6.8)–(6.10)],

$$
\begin{aligned}
P_n(\varepsilon_1) &= O\big(((n-1)!)^{-1}\exp(-n^{\sigma})\big),\\
\forall\,\sigma &< \sigma(\varepsilon_1) = 1-(1-\varepsilon_1)^{1/2}.
\end{aligned}
\tag{6.11}
$$

Therefore, by (6.6),

$$
\pi_n(\varepsilon_1) = O(\exp(-n^{\sigma})), \qquad \forall\,\sigma<\sigma(\varepsilon_1).
\tag{6.12}
$$

Consequently, with probability at least $1-O(\exp(-n^{\sigma}))$, each stable matching contains at least one pair (m_i, w_j) such that

$$
Y_{ij} \ge (1-\varepsilon_1)\frac{\log^2 n}{n}.
$$

STEP 2. Let $\varepsilon\in(\varepsilon_1, 1)$. By N_j we denote the total number of entries in the jth column of Y which are less than $(1-\varepsilon_1)(\log^2 n)/n$, $1\le j\le n$. Each N_j is binomially distributed with parameters n and $p=(1-\varepsilon_1)(\log^2 n)/n$ [$N_j = \mathrm{Bin}(n,p)$, in short]. Using the Chernoff inequality [2],

$$
\begin{aligned}
P(\mathrm{Bin}(n,p)\le k) &\le \exp\bigg(nH\Big(\frac{k}{n}\Big)\bigg), \qquad \forall\,k\le np,\\
H(x) &= x\log\frac{p}{x} + (1-x)\log\frac{1-p}{1-x},
\end{aligned}
\tag{6.13}
$$

we obtain easily

$$
P\big(N_j\le(1-\varepsilon)\log^2 n\big) = O\big(\exp(-(\log^2 n)h(\varepsilon,\varepsilon_1))\big),
$$

where

$$
h(\varepsilon,\varepsilon_1) =_{\text{def}} \varepsilon-\varepsilon_1+(1-\varepsilon)\log\frac{1-\varepsilon}{1-\varepsilon_1} > 0
$$

for each $\varepsilon_1\in[0,\varepsilon)$. Therefore, each column of Y contains *at least* $(1-\varepsilon)\log^2 n$ entries less than $(1-\varepsilon_1)(\log^2 n)/n$, with probability at least $1-O(\exp(-(\log^2 n)h_1))$, $\forall\, h_1 < h(\varepsilon,\varepsilon_1)$.

Combining the conclusions of Steps 1 and 2, we are able to state that

$$P(\mathscr{R} \geq (1-\varepsilon)\log^2 n) \geq 1 - O\big(\max\big(\exp(-n^{\sigma}), \exp(-(\log^2 n)h_1)\big)\big)$$
$$= 1 - O\big(\exp(-(\log^2 n)h_1)\big).$$

So, the relation (6.1) for $\mathscr{R}$ (whence for $\mathscr{Q}$) follows by letting $\varepsilon_1 \downarrow 0$, and noticing that $h(\varepsilon, 0) = c(\varepsilon)$ which is defined in (6.2).

(b) To derive the upper bound (6.3) (again, for $\mathscr{R}$ instead of $\mathscr{Q}$), we write first

(6.14) $$\{\mathscr{R} \geq (2+a)\log^2 n\} \subseteq A_1 \cup A_2;$$

here

$$A_1 = \left\{\mathscr{R} \geq (2+a)\log^2 n \text{ and } Q(\mathscr{M}_2) \geq \frac{n^2}{(1+\lambda)\log n}\right\},$$

$$A_2 = \left\{Q(\mathscr{M}_2) < \frac{n^2}{(1+\lambda)\log n}\right\}.$$

[Bringing $\mathscr{R}$ and $Q(\mathscr{M}_2)$ together is not unnatural since $\mathscr{R} = \max_j R_j(\mathscr{M}_2)$.] Using (5.15), we have

(6.15) $$P(A_2) = O(n^{-\lambda'}), \qquad \forall\, \lambda' < \lambda.$$

Let us estimate $P(A_1)$. To this end, define B as the event that in $\mathscr{M}_2$ there is a pair (m_i, w_j) such that

(6.16) $$Y_{ij} \geq \alpha =_{\text{def}} (2 + a_1)\frac{\log^2 n}{n},$$

and

$$Q(\mathscr{M}_2) \geq \frac{n^2}{(1+\lambda)\log n}.$$

Here $a_1 \in (0, a)$ is fixed. Using (3.3) and the Chernoff method (cf. the proof of Theorem 4.1), we obtain

(6.17) $$P(B) \leq nn! \int_{C(\alpha)} \inf\{F(\xi, x, y) \colon \xi \geq 1\}\, dx\, dy,$$

where

$$C(\alpha) = \{(x,y) \colon 0 \leq x_i, y_j \leq 1\ (1 \leq i, j \leq n) \text{ and } y_1 \geq \alpha\},$$

(6.18) $$F(\xi, x, y) = \xi^{n-k} \prod_{i \neq j} (\bar{x}_i + x_i \bar{y}_j \xi)$$

and

(6.19) $$k = \frac{n^2}{(1+\lambda)\log n}.$$

[The integral is an upper bound for the probability that the standard matching

M_0 is stable, $Q(M_0) \geq k$ and $Y_{11} \geq \alpha$.] Using (6.17), we will find an explicit bound for $P(B)$, and that will enable us to bound $P(A_1)$, too.

Let us estimate first the contribution P_1 to the bound in (6.17) which corresponds to

$$C_1(\alpha) = \left\{(x,y) \in C(\alpha) \colon s\left(= \sum_{i=1}^{n} x_i\right) \leq \tfrac{1}{2}\log n\right\}.$$

For this purpose, we set $\xi \equiv 1$ in (6.17) and (6.18) and, like the bound (6.8), we obtain [disregarding the condition $y_1 \geq \alpha$ and dropping the $P(\max_i \mathscr{L}_i \leq s^{-1})$ factor]

$$\begin{aligned} P_1 &= O\left(nn! \int_{s \leq (1/2)\log n} \Psi_1^n(s)\, ds\right) \qquad \left(\Psi_1(s) = \frac{1 - e^{-s}}{s}\right) \\ &= O\left(n^2 \log n (1 - n^{-1/2})^n\right) \\ &= O\left(\exp(-n^{1/3})\right). \end{aligned} \tag{6.20}$$

For the domain $C_0(\alpha) = C(\alpha) \setminus C_1(\alpha)$, we set $t = k/n$ and define

$$\xi = \xi(s) = \begin{cases} t/s, & \text{for } s < t, \\ 1, & \text{for } s \geq t. \end{cases}$$

To estimate P_0, the contribution of $C_0(\alpha)$, we bound each factor $(\bar{x}_i + x_i \bar{y}_j \xi)$ in F by $\exp(x_i(\xi - 1) - x_i y_j \xi)$ and integrate over y again, keeping the restriction $y_1 \geq \alpha$ this time:

$$\begin{aligned} P_0 \leq nn! \int_{C_0(\alpha)} & \xi^{n-k} \exp(s(n-1)(\xi - 1)) \\ & \times \prod_{j \neq 1} \Psi_1(s_j \xi) \left(\frac{\exp(-\alpha s_1 \xi) - \exp(-s_1 \xi)}{s_1 \xi}\right) dx \end{aligned} \tag{6.21}$$

($s_j =_{\text{def}} \sum_{i \neq j} x_i$). On $C_0(\alpha)$,

$$\frac{s_j}{s} = 1 + O(\log^{-1} n);$$

using inequality (5.6) again, we obtain

$$\begin{aligned} \prod_{j \neq 1} \Psi_1(s_j \xi) &\leq \Psi_1^{n-1}(s\xi) \exp\left(\frac{c}{\xi \min s_j + 1} \sum_{i \neq 1} x_i \xi\right) \\ &= O\left(\Psi_1^{n-1}(s\xi)\right). \end{aligned}$$

Besides,

$$\frac{\exp(-\alpha_1 s_1 \xi) - \exp(-s_1 \xi)}{s_1 \xi} = O\left(\frac{\exp(-\alpha_1 s \xi)}{s \xi}\right),$$

where

$$\alpha_1 = \alpha\left(1 + O(\log^{-1} n)\right). \tag{6.22}$$

So, replacing $C_0(\alpha)$ in (6.21) by a larger domain $\{x \geq 0: s \leq n\}$ and turning to integration over s, we have [similarly to (4.19)]

$$P_0 = O\left(n^2 \int_0^n e^{G(s)}\, ds\right),$$

where

$$G(s) = (n-k)\log \xi + s(n-1)(\xi - 1) - n\log(s\xi) - \alpha_1 s\xi + (n-1)\log s.$$

(i) Suppose $s \geq t$. Then $\xi \equiv 1$ and

$$G(s) = -\log s - \alpha_1 s,$$

that is,

$$n^2 \int_t^n e^{G(s)}\, ds \leq \frac{n^2}{t} \int_t^\infty e^{-\alpha_1 s}\, ds = \frac{n^2 e^{-\alpha_1 t}}{\alpha_1 t} = O(n^{-f}),$$

(6.23) $$f = \frac{2 + a_1}{1 + \lambda} - 2;$$

$t = k/n$ [see also (6.16) and (6.19)].

(ii) Suppose $s < t$. Then $\xi = t/s$ and, after cancellations,

$$G(s) = k \log\frac{s}{t} - \log s + (n-1)(t-s) - \alpha_1 t.$$

Now

$$G''(s) = -\frac{k-1}{s^2} < 0,$$

so G is concave down, and G reaches its maximum at

$$t_0 = \frac{k-1}{n-1} = t + O\left(\frac{k}{n^2}\right).$$

Hence,

$$G(t_0) = G(t) + O\left(\frac{k}{n^2}\right),$$

and, appealing to Laplace's method of estimating integrals,

$$n^2 \int_0^t e^{G(s)}\, ds = O\left(n^2 e^{G(t_0)} \left(-G''(t_0)\right)^{1/2}\right) = O(n^{-f}),$$

[see (6.23) for f].

Therefore, $P_0 = O(n^{-f})$, whence by (6.20),

(6.24) $$P(B) = O(n^{-f}).$$

By $\mathcal{N}_j$ we denote the total number of entries in the jth column of Y which are less than $(2 + a_1)(\log^2 n)/n$ (compare with definition of N_j in Step 2). Using the Chernoff bound (6.13) again, we obtain, for $a > a_1$,

$$(6.25) \qquad P\big(\mathcal{N}_j \geq (2+a)\log^2 n\big) = O\big(\exp(-(\log^2 n) g(a, a_1))\big),$$

where

$$g(a, a_1) =_{\text{def}} (2+a)\log \frac{2+a}{2+a_1} + a_1 - a < 0.$$

Introduce $A_3 = \bigcup_j \{\mathcal{N}_j \geq (2+a)\log^2 n\}$, the event that at least one column of Y contains at least $(2+a)\log^2 n$ entries less than $(2+a_1)(\log^2 n)/n$. The estimate (6.25) implies that

$$(6.26) \qquad P(A_3) = O\big(\exp(-(\log^2 n) g_1)\big), \qquad \forall\, g_1 < g(a, a_1).$$

Then, recalling the definition of the events A_1 and B,

$$\begin{aligned} P(A_1) &\leq P(B) + P(A_1 B^c) \\ &\leq P(B) + P(A_3). \end{aligned}$$

Hence, by (6.24) and (6.26), for each $a_1 < a$ we have

$$P(A_1) = O(n^{-f}), \qquad f = \frac{2+a_1}{1+\lambda} - 2,$$

and [see (6.14) and (6.15)]

$$P(\mathcal{R} \geq (2+a)\log^2 n) = O(n^{-\lambda'} + n^{-f}), \qquad \forall\, \lambda' < \lambda.$$

Thus,

$$P(\mathcal{R} \geq (2+a)\log^2 n) = O(n^{-c}), \qquad \forall\, c < c(a),$$

where

$$\begin{aligned} c(a) &= \max\left\{\min\left(\lambda, \frac{2+a}{1+\lambda} - 2\right) : \lambda \geq 0\right\} \\ &= 2a\left[3 + (4a+9)^{1/2}\right]^{-1}. \end{aligned}$$

The proof of Theorem 6.1 is complete. □

The next theorem reveals that, in a sharp contrast, $\max_j R_j(\mathcal{M}_1)$ and $\max_i Q_i(\mathcal{M}_2)$ are whp asymptotic to n. [Needless to say, $\max_j R_j(\mathcal{M}_1)$ and $\max_i Q_i(\mathcal{M}_2)$ are equidistributed.] Thus, the worst husband in $\mathcal{M}_1$ (the worst wife in $\mathcal{M}_2$) is almost at the bottom of his wife's (her husband's) order list of men (women)!

THEOREM 6.2. *For every* $\delta \in (0, 1)$

$$P\Big(\max_j R_j(\mathcal{M}_1) \geq (1-\delta)n\Big) \geq 1 - O(n^{-c}), \qquad \forall\, c < c(\delta),$$

where

$$c(\delta) = \frac{\delta}{1-\delta}.$$

PROOF. Fix $c < c(\delta)$ and write

$$P\Big(\max_j R_j(\mathscr{M}_1) \le (1-\delta)n\Big) \le P(A_4) + P(A_5),$$

where

$$A_4 = \Big\{\max_j R_j(\mathscr{M}_1) \le (1-\delta)n \text{ and } Q(\mathscr{M}_1) \le (1+c)n\log n\Big\},$$

$$A_5 = \{Q(\mathscr{M}_1) > (1+c)n\log n\}.$$

Then [see (5.15)]

$$P(A_5) = O(n^{-c}).$$

To estimate $P(A_4)$, introduce $\delta_1 \in (0, \delta)$ so close to δ that

(6.27)
$$c < c(\delta_1) = \frac{\delta_1}{1-\delta_1},$$

and define B_1 the event that for each pair (m_i, w_j) in $\mathscr{M}_1$,

$$Y_{ij} \le 1 - \delta_1$$

and

$$Q(\mathscr{M}_1) \le (1+c)n\log n.$$

Analogously to (6.17) and (6.18),

$$P(B_1) \le n!\int_{C(\delta)} \inf\{F(\xi,x,y) : \xi \le 1\}\,dx\,dy,$$

where $k = (1+c)n\log n$ and

$$C(\delta) = \big\{(x,y) \in C : y_j \le 1-\delta_1, 1 \le j \le n\big\}.$$

Following the established pattern, we have eventually

$$P(B_1) = O\bigg(n\int_0^n e^{G_1(s)}\,ds\bigg).$$

Here

$$G_1(s) = (n-1)(\xi-1)s + n\log\frac{1-\exp(-(1-\delta_1)s\xi)}{s\xi}$$
$$+(n-k)\log\xi + (n-1)\log s,$$

and

$$\xi = \xi(s) = \begin{cases} 1, & s < t, \\ \dfrac{t}{s}, & s \geq t, \end{cases} \qquad \left(t =_{\text{def}} \frac{k}{n} \right).$$

Using condition (6.27) [$\Leftrightarrow (1 + c)(1 - \delta_1) < 1$], it is easy to show that

$$\begin{aligned} \{\max G_1(s) \colon 0 \leq s \leq n\} &= G_1(t) + O(\log n/n) \\ &= -n^{1-(1-\delta_1)(1+c)}(1 + o(1)). \end{aligned}$$

So, $P(B_1)$ is superpolynomially small.

Also quite small (indeed, exponentially small) is $P(A_6)$, where A_6 is the event that some column of Y has at least $n\delta$ entries exceeding $n(1 - \delta_1)$. [Predictably, the proof uses the Chernoff bound.]

Then

$$\begin{aligned} P(A_4) &\leq P(B_1) + P(A_4 B_1^c) \\ &\leq P(B_1) + P(A_6) \\ &= O(P(B_1)). \end{aligned}$$

Therefore,

$$\begin{aligned} P\Big(\max_j R_j(\mathscr{M}_1)\Big) &= O(P(B_1) + P(A_5)) \\ &= O(n^{-c}) \end{aligned}$$

for every $c < c(\delta)$, which completes the proof. □

Our next (and last) theorem shows that the egalitarian matching $\mathscr{M}_3$ fares much better than either $\mathscr{M}_1$ or $\mathscr{M}_2$, in the sense that no spouse in $\mathscr{M}_3$ has a rank which noticeably exceeds $n^{1/2}$. In fact, whp $\mathscr{M}_3$ is "optimal," that is, there does not exist a stable matching $\mathscr{M}$ in which the largest rank of a spouse falls significantly below $n^{1/2}$. Here is the precise statement.

THEOREM 6.3. *Define* $\mathscr{S}(\mathscr{M}) = \max(\mathscr{Q}(\mathscr{M}), \mathscr{R}(\mathscr{M}))$, *that is*, $\mathscr{S}(\mathscr{M})$ *is the largest rank of a spouse in a stable matching* $\mathscr{M}$.

(a) *For every* $b > 0$ *and* $c < b$,

$$P\big(\mathscr{S}(\mathscr{M}_3) \leq (\tfrac{7}{4} + b)n^{1/2} \log n\big) \geq 1 - O(n^{-c}). \tag{6.28}$$

(b) *For each* $\varepsilon \in (0, 1)$ *and* $\sigma < \min(\varepsilon, \frac{1}{2})$,

$$P\Big(\min_{\mathscr{M}} \mathscr{S}(\mathscr{M}) \geq (1 - \varepsilon)n^{1/2} \log n\Big) \geq 1 - O(\exp(-n^{\sigma})). \tag{6.29}$$

PROOF. (a) Let A_7 (A_8) be the event that in $\mathscr{M}_3$ there is a pair (m_i, w_j) such that $Y_{ij} \geq \Delta$ ($X_{ij} \geq \Delta$, respectively); $\Delta = (\frac{7}{4} + b)\log n/n^{1/2}$. By symmetry, $P(A_7) = P(A_8)$. If we show that, for every $b > 0$, $P(A_7) = O(n^{-c})$, $\forall\ c < b$, then the bound (6.28) will follow via the Chernoff estimates, in a manner we have seen twice before.

Using the bound (5.17), we can write

$$P(A_7) \le nn! \int_{C(\Delta)} \inf\{\Phi(\xi,\eta,x,y): \xi,\eta \in (0,1]\}\,dx\,dy + O(n^{-a}), \quad \forall a > 0, \tag{6.30}$$

where Φ is as defined in (4.9),

$$k = l = (1 + n^{-\sigma})n^{3/2}, \qquad \sigma \in (0, \tfrac{1}{4}), \tag{6.31}$$

and $C(\Delta)$ is the part of C on which $y_1 \ge \Delta$. The intuition behind (6.30) is that with very high probability $R(\mathscr{M}_3)$ and $Q(\mathscr{M}_3)$ are *both* at most $(1 + n^{-\sigma})n^{3/2}$, and under this condition it should be unlikely that, for $(m_i, w_j) \in \mathscr{M}_3$, Y_{ij} is considerably larger than $n^{-1/2}$. To bound sharply the integral in (6.30), we set

$$\xi = \xi(s) = \begin{cases} 1, & \text{if } s < t_1, \\ \dfrac{t_1}{s}, & \text{if } s \ge t_1, \end{cases} \qquad \eta = \eta(s) = \begin{cases} \dfrac{n-1-t_2}{n-1-s}, & s < t_2, \\ 1, & s \ge t_2, \end{cases} \tag{6.32}$$

where $t_1 = k/n$, $t_2 = n^2/l$. [As it should be, $\xi(s)$, $\eta(s) \in (0,1]$ for each s.] By the condition (6.31), $t_1 > t_2$. Thus, in a nice contrast with $\xi^{\pm}(s)$, $\eta^{\pm}(s)$ (see the proof of Theorem 4.1), there is no need for a breakup point between t_2 and t_1.

Appealing to (4.17), we have [cf. (4.18)]

$$\Phi(\xi,\eta,x,y) \le \xi^{n-k}\eta^{n-l}\exp\left[s(n-1)(\xi-1) - \sum_j s_j(\xi,\eta)y_j\right],$$

where

$$s_j(\xi,\eta) = s(\xi,\eta) - (\xi + \eta - 1)x_j,$$

and

$$s(\xi,\eta) = s\xi - (n-1-s)(\eta-1).$$

In particular,

$$s(\xi,\eta) = \begin{cases} t_2, & s < t_2, \\ s, & t_2 \le s < t_1, \\ t_1, & s \ge t_1; \end{cases} \tag{6.33}$$

that is, $s(\xi,\eta)$ is of order precisely $n^{1/2}$, uniformly over $0 \le s \le n$. Besides, since ξ, η, $x_j \in [0,1]$,

$$s_j(\xi,\eta) = s(\xi,\eta) + O(1).$$

Using inequality (5.6) and

$$\int_\Delta^1 \exp(-s_1(\xi,\eta)y)\,dy = (1 + o(1))\frac{\exp(-s(\xi,\eta)\Delta)}{s(\xi,\eta)},$$

we obtain

$$P(A_7) = O\left(n^2 \int_0^n e^{G_2(s)}\, ds\right).$$

Here

$$G_2(s) = (n-k)\log \xi + (n-l)\log \eta + s(n-1)(\xi - 1)$$
$$- n \log s(\xi, \eta) - s(\xi, \eta)\Delta + (n-1)\log s.$$

It can be checked easily, via (6.31)–(6.33), that

$$n^2 \int_{t_2}^{t_1} e^{G_2(s)}\, ds = O\left(n^2 \int_{t_1}^{t_2} \frac{e^{-s\Delta}}{s}\, ds\right)$$
$$= O(n^{-(b+\sigma-1/4)}),$$
$$n^2 \int_0^{t_2} e^{G_2(s)}\, ds = O(n^2 e^{G_2(t_2)})$$
$$= O(n^{-(b+1/4)})$$
$$n^2 \int_{t_1}^{n} e^{G_2(s)}\, ds = O(n^{-(b+1/2)}).$$

Thus,

$$P(A_7) = O(n^{-(b+\sigma-1/4)}), \qquad \forall\, \sigma < \tfrac{1}{4},$$

that is,

$$P(A_7) = O(n^{-c}), \qquad \forall\, c < b.$$

The proof of (6.28) is complete.

(a) To prove (6.29), introduce the event A_9 that for some stable matching $\mathscr{M}$ and every pair (m_i, w_j) in it,

$$X_{ij} \le \Delta_1, \qquad Y_{ij} \le \Delta_1,$$

where

$$\Delta_1 = (1 - \varepsilon_1)\frac{\log n}{n^{1/2}}, \qquad \varepsilon_1 \in (0, \varepsilon).$$

Then (cf. Step 1 in the proof of Theorem 6.1),

$$P(A_9) \le n! \int_{x_i, y_i \le \Delta_1} \prod_{i \ne j} (1 - x_i y_i)\, dx\, dy$$
$$= n! \Delta_1^{2n} \int_C \prod_{i \ne j} \left(1 - \Delta_1^2 x_i y_j\right) dx\, dy$$
$$= O\left(n \int_0^n s^{-1}\left(1 - \exp(-s\,\Delta_1^2)\right)^n P\left(\max_{1 \le i \le n} \mathscr{L}_i \le s^{-1}\right) ds\right).$$

Fix $\lambda \in ((1-\varepsilon_1)^2, 1)$ and break up the interval $[0, n]$ into $[0, t]$ and $[t, n]$, where $t = n/(\lambda \log n)$. Then, dropping the probability $P(\max_{1 \le i \le n} \mathscr{L}_i \le s^{-1})$,

$$\int_0^t = O\Big(t\big(1 - \exp(-t\Delta_1^2)\big)^{n-1}\Big) = O(\exp(-n^{\mu_1})), \qquad \forall\, \mu_1 < 1 - (1-\varepsilon_1)^2/\lambda.$$

Also, using the bound (6.10) for this probability at $s = t$ and dropping $(1 - \exp(-s\,\Delta_1^2))^n$,

$$\int_t^n = O\left(\frac{n-t}{t}\exp(-n^{\mu_2})\right), \qquad \forall\, \mu_2 < 1 - \lambda.$$

For $\lambda = 1 - \varepsilon_1$ the bounds of μ_1 and μ_2 are both equal to ε_1, and we conclude that

$$P(A_9) = O(\exp(-n^{\gamma})), \qquad \forall\, \gamma < \varepsilon_1. \tag{6.34}$$

Furthermore, denoting the event in (6.29) by A_{10}, we have

$$P(A_{10}^c) \le P(A_9) + P(A_9^c A_{10}^c). \tag{6.35}$$

We can see that on the event $A_9^c A_{10}^c$ there is a row in X, or a column in Y, such that the total number of entries in this row (column) which do not exceed Δ_1 is at most $(1-\varepsilon)n^{1/2}\log n$. Sure enough, the Chernoff bound (6.13) shows then that

$$P(A_9^c A_{10}^c) = O\big(\exp(-\rho n^{1/2}\log n)\big), \qquad \rho < \rho(\varepsilon_1, \varepsilon), \tag{6.36}$$

where

$$\rho(\varepsilon_1, \varepsilon) = \varepsilon - \varepsilon_1 + (1-\varepsilon)\log\frac{1-\varepsilon}{1-\varepsilon_1} > 0,$$

for $\varepsilon_1 < \varepsilon$. Hence,

$$P(A_{10}^c) = O(\exp(-n^{\sigma})), \qquad \forall\, \sigma < \min(\varepsilon, \tfrac{1}{2}).$$

The proof of (6.29) and thus of Theorem 6.3 is complete. □

NOTE 1. This result shows that whp $\min_{\mathscr{M}} \mathscr{S}(\mathscr{M})$ is sandwiched between $(1-\varepsilon)n^{1/2}\log n$ and $(\frac{7}{4} + \varepsilon)n^{1/2}\log n$. Forced to make a guess, we would conjecture that, out of these bounds, the upper bound is sharper. (*See the Note added in proof at the end of the text.*)

NOTE 2. How small is the *smallest* rank of a spouse in a stable matching? Let K_n be the total number of perfect matches, that is, the pairs (m_i, w_j) such that m_i and w_j are the first choice partners for each other. Obviously, for every such pair, m_i and w_j must be matched in every stable matching. It is clear that the rth factorial moment of K_n is given by

$$E(K_n^{\underline{r}}) = (n^{\underline{r}})^2/n^{2r}, \qquad r \ge 1,$$

where $a^{\underline{b}} =_{\text{def}} a(a-1)\cdots(a-b+1)$. Therefore $E(K_n^{\underline{r}}) \to 1$ as $n \to \infty$, for each $r \geq 1$, and consequently

$$P(K_n = r) \to \frac{e^{-1}}{r!}, \qquad r \geq 0,$$

that is, K_n is asymptotically Poisson with parameter 1. In particular, with the limiting probability greater than or equal to $1 - e^{-1}$, the smallest rank of a spouse (in every stable matching) is 1. In fact, this limiting probability *equals* 1. Here is how it can be proven. Denote by T_n the total number of women whose husbands in the male-optimal stable matching $\mathscr{M}_1$ happen to be their best choices. In the notation of Section 2, let $V_{n1}(w_j)$ stand for the total number of proposals received by the woman w_j in the first phase of the proposal algorithm (i.e., the McVitie–Wilson algorithm). Conditioned on the values of $V_{n1}(w_j)$, $1 \leq j \leq n$, the events $\mathscr{A}_j = \{w_j\text{'s husband in } \mathscr{M}_1 \text{ has rank } 1\}$ are *independent*, with the conditional probabilities $P(\mathscr{A}_j | \cdot)$ given by

$$\begin{aligned} P(\mathscr{A}_j | \cdot) &= \binom{n}{V_{n1}(w_j) - 1} \Big/ \binom{n}{V_{n1}(w_j)} \\ &= \frac{V_{n1}(w_j)}{n}. \end{aligned}$$

We proved in [21] that $\sum_j V_{n1}(w_j)$ the running time of the first phase whp satisfies

$$\sum_j V_{n1}(w_j) = n(\log n + O(\log\log n)). \tag{6.37}$$

It is not difficult to show, using (6.37), that whp

$$\max_j V_{n1}(w_j) \leq e \log n. \tag{6.38}$$

But if $V_{n1}(w_j)$, $1 \leq j \leq n$, satisfy (6.37) and (6.38) then the *conditional* distribution of T_n (the sum of set indicators for the events $\mathscr{A}_j$) is asymptotically normal with mean and variance both equal to $\log n + O(\log\log n)$. Since $\log\log n = o(\log^{1/2} n)$, we obtain that, unconditionally, $(T_n - \log n)/\log^{1/2} n \Rightarrow \mathscr{N}(0, 1)$. In particular, $P(T_n > 0) \to 1$ as $n \to \infty$.

An interesting open problem is the determination of the likely range of T_n', the total number of women each of whom has precisely one stable husband. Of course, $T_n' \geq T_n$ and the question is how large $T_n' - T_n$ is typically.

7. Three corollaries. Let us have a closer look at the *first phase* of the probabilistic algorithm described in the proof of Theorem 2.1. Following Wilson [23], let us slightly modify the algorithm. Each man who is asked to propose proposes to a woman chosen uniformly at random among all women, including those who have rejected him earlier. If a man does make a repeated proposal to a woman, then his proposal is rejected, which is natural since at this time the woman certainly holds a better proposal. Knuth [17] called these two categories of proposals nonredundant and redundant, respectively, and he interpreted the men's behavior in terms of "amnesia," that is, inability of the men to keep records of their past proposals. Deleting the redundant proposals, we obtain a sequence of (nonredundant) proposals, which has exactly the same distribution as the sequence generated by the work of the original algorithm. In particular, $N \equiv_{\mathscr{D}} Q(\mathscr{M}_1)$, where N is the total number of nonredundant proposals and $Q(\mathscr{M}_1)$ is the total wives' rank in the stable matching $\mathscr{M}_1$. Further, if two consecutive nonredundant proposals P_{s-1} and P_s, $2 \le s \le Q(\mathscr{M}_1)$, are separated by a run of r_s redundant proposals, then all of the latter are by a man m who makes the proposal P_s. Let the number of different women whom the man m has asked by the moment immediately after the proposal P_{s-1} be equal to k. Then, *conditioned* on the state of the process at this moment,

$$P(r_s = j \mid \cdot) = pq^j, \qquad j \ge 0,$$

where

$$p = 1 - k/n, \qquad q = 1 - p.$$

Introduce ρ, the total number of redundant proposals; clearly,

$$\rho = \sum_{s=1}^{Q(\mathscr{M}_1)} r_s. \tag{7.1}$$

Define also ρ_i and ρ_{ij} as the total number of redundant proposals made by the man m_i by the end of the process and between his $(j-1)$th and jth nonredundant proposals, respectively. Like r_s, ρ_{ij} is geometrically distributed with parameter $p = 1 - (j-1)/n$. Rearranging terms in (7.1), we have

$$\rho = \sum_{i=1}^{n} \sum_{j=2}^{Q_i(\mathscr{M}_1)} \rho_{ij}, \tag{7.2}$$

where $Q_i(\mathscr{M}_1)$ is the total number of nonredundant proposals by the man m_i, in other words, the rank of his wife in $\mathscr{M}_1$.

Wilson observed that in the extended sequence of proposals each proposal is made to a woman selected uniformly at random, regardless of who proposes. Therefore, $N + \rho \equiv_{\mathscr{D}} D$, where D is the total number of draws in a classic coupon-collector problem, whence the upper bound

$$E(Q(\mathscr{M}_1)) = E(N) \le E(D) = nH_n.$$

[In fact, our upper bound (5.15) for $Q(\mathscr{M}_1)$ and $R(\mathscr{M}_2)$ is based on $N \le D$ and an inequality

$$P(D_n \le (1+a)n \log n) \ge 1 - n^{-a}; \tag{7.3}$$

see [22].]

Knuth [17] found a better bound,

$$E(Q(\mathscr{M}_1)) \le (n-1)H_n + 1, \tag{7.4}$$

by considering a case of *partial amnesia*, when a rejected man proposes to a woman selected among all women except the one who has just rejected him.

Knuth also was able to show that

$$E(\rho) = O(\log^4 n),$$

thereby establishing a lower bound,

$$E(Q(\mathscr{M}_1)) \geq nH_n - O(\log^4 n).\ (!)$$

The following is a sharper lower bound based on Theorem 6.1 of this paper.

COROLLARY 7.1.

$$E(Q(\mathscr{M}_1)) \geq nH_n - (2 + o(1))\log^3 n. \tag{7.5}$$

PROOF. The proof follows from two observations. First, by (7.1),

$$E\Big(\rho \big| Q(\mathscr{M}_1) = l, \mathscr{Q} = \max_i Q_i(\mathscr{M}_1) = k\Big) \leq lE(r(k)) = l\frac{k/n}{1 - k/n},$$

where $r(k)$ is geometrically distributed with parameter $p = 1 - k/n$. Second, according to (5.15) and (6.3), for every $\varepsilon > 0$,

$$\begin{aligned} &P\big(Q(\mathscr{M}_1) \leq (1+\varepsilon)n \log n\big) \geq 1 - n^{-\varepsilon}, \\ &P\big(\mathscr{Q} \leq (2 + \varepsilon_1)\log^2 n\big) \geq 1 - n^{-\varepsilon}, \qquad \forall\, \varepsilon_1 > \varepsilon. \end{aligned} \tag{7.6}$$

Then, since $Q(\mathscr{M}_1)\mathscr{Q} \leq n^3$ always,

$$\begin{aligned} E(\rho) &\leq (1 + o(1))n \log n(n^{-1} 2 \log^2 n) \\ &= (1 + o(1)) 2 \log^3 n. \end{aligned}$$

□

Curiously, the identity (7.2) is useful, too. We have

$$\begin{aligned} E(\rho) &= nE\left(\sum_{j=2}^{Q_1(\mathscr{M}_1)} \frac{(j-1)/n}{1 - (j-1)/n}\right) \geq \frac{1}{2} E(Q_1(\mathscr{M}_1)(Q_1(\mathscr{M}_1) - 1)) \\ &\geq \left(\frac{1}{2} + o(1)\right) E^2(Q_1(\mathscr{M}_1)) \\ &= \left(\frac{1}{2} + o(1)\right) \log^2 n, \end{aligned}$$

whence

$$E(Q(\mathscr{M}_1)) \leq nH_n - \left(\tfrac{1}{2} + o(1)\right)\log^2 n, \tag{7.7}$$

which improves Knuth's bound (7.4).

Notice that, by (7.5) and the formula for $E(\rho)$, $E(\rho)$ is actually asymptotic to $\frac{1}{2}E(Q_1^2(\mathscr{M}_1))$. Is $\log^2 n$ a correct order of $E(Q_1^2(\mathscr{M}_1))$ and thus of $E(\rho)$? After all, $E(Q_1(\mathscr{M}_1)) \sim \log n$, and it would seem plausible that the likely values of $Q_1(\mathscr{M}_1)$ are essentially of logarithmic order. However, $E(\max_i Q_i(\mathscr{M}_1))$ is of order $\log^2 n$, so we are inclined toward $E(\rho)$ being of order $(\log\log n)\log^2 n$, say. It would be very interesting to resolve the question one way or another.

Turn now to a lower bound problem for S_n (the total number of stable matchings). By Theorem 6.1, the largest rank of a husband in $\mathscr{M}_2$ is q.s. of order $\log^2 n$. However, most of the husbands in $\mathscr{M}_1$ q.s. have ranks at least $\log^3 n$, say. Here is a proof.

Let $V_{n1}(w_j)$ stand for the total number of the nonredundant proposals made to the woman w_j. Suppose $a > a' > 0$. On the event $\{D \le (1 + a')n \log n\}$, $V_{n1}(w_j)$ is stochastically dominated by Bin(ν, p), with $\nu = (1 + a')n \log n$ and $p = n^{-1}$. So, using (7.3) and Chernoff's bound,

$$P\Big(\max_j V_{n1}(w_j) \ge (1 + a)\log n\Big) = O(n(n^{-a'} + n^{-a''})),$$

$$a'' = (1 + a)\log\frac{1 + a}{1 + a'} + a' - a.$$

Selecting $a' < a$ so as to equate a' and a'', we get

$$P\Big(\max_j V_{n1}(w_j) \ge (1 + a)\log n\Big) = O(n^{-b}), \tag{7.8}$$

where

$$b = (1 + a)\exp\Big(-\frac{a}{1 + a}\Big) - 2 \ge (1 + a)e^{-1} - 2, \tag{7.9}$$

that is, b is positive for $a > 2e - 1$. Further, given the values $v_j = V_{nj}$, the random variables $R_j(\mathscr{M}_1)$, $1 \le j \le n$ (the ranks of the husbands in $\mathscr{M}_1$), are *independent*, and [see (2.2) and (2.3)]

$$P\big(R_j(\mathscr{M}_1) \le \log^3 n \mid \cdot\big) = 1 - \binom{n - \log^3 n}{v_j}\Big/\binom{n}{v_j}$$
$$\le v_j(\log^3 n)/(n - v_j),$$

that is, at most $(2 + a)(\log^4 n)/n$ if

$$\max_j v_j \le (1 + a)\log n. \tag{7.10}$$

Thus, in the standard way,

$$W =_{\text{def}} \big|\{w_j\colon R_j(\mathscr{M}_1) \le \log^3 n\}\big| \le (3 + a)\log^4 n$$

with the conditional probability greater than or equal to $1 - \exp(-c(a)\log^4 n)$, $c(a) > 0$, uniformly over $\{v_j\colon 1 \le j \le n\}$ satisfying (7.10). Hence, the unconditional probability of the event $\{W \le (3 + a)\log^4 n\}$ is at least $1 - O(n^{-b})$ [see (7.8) and (7.9)], which ends the proof.

Consequently, with probability greater than or equal to $1 - O(n^{-b})$, the rotations in a chain of stable matchings, which connects $\mathscr{M}_1$ and $\mathscr{M}_2$, must overall involve at least $n - (3 + a)\log^4 n$ women; [see (7.9) for b]. On the other hand (Theorem 5.1), with probability greater than or equal to $1 - O(n^{-\alpha})$, no rotation may involve more than $((1 + a)n \log n)^{1/2}$ women. Combination of these statements produces the following result.

COROLLARY 7.2. *For every* $\alpha > 0$,

$$P\left(S_n \geq \sqrt{\frac{n}{(1+\alpha)\log n}}\right) \geq 1 - O(n^{-\alpha'}), \qquad \forall\, \alpha' < \alpha.$$

(S_n is at least as large as the length of a chain between $\mathscr{M}_1$ and $\mathscr{M}_2$.)

Our next (and last) statement concerns a *constrained* stable matching problem. It is assumed that for each man m_i there is a set W_i of women he may propose to. In this case, a stable matching does not always exist. Still, whenever a stable matching exists, the McVitie–Wilson algorithm will find a (male-optimal) stable matching.

Suppose that $|W_1| = \cdots = |W_n| = k$. We postulate that the random instance of ranking is obtained as follows. Each man m_i chooses, independently of other men and uniformly at random, an ordered sequence $W_i = (w_{j_1}, \dots, w_{j_k})$ of k women, and each woman w_j ranks, uniformly at random, all the men m_i such that $w_j \in W_i$. Appealing to the principle of deferred decisions, we can avoid generating the full ranking system in advance, unfolding it instead via the steps of the probabilistic proposal algorithm. Thus, the algorithm will result in a (male-optimal) stable matching iff each of n women receives a proposal before one of the men gets k rejections.

Compared to n, how large must k be to guarantee that, with high probability as $n \to \infty$, the random instance has (does not have) a stable matching? In the light of the preceding discussion, the answer is contained in Theorem 6.1.

COROLLARY 7.3. (a) *If* $k \leq (1-\varepsilon)\log^2 n$, $\varepsilon \in (0,1)$, *then no stable matching exists with probability greater than or equal to* $1 - \exp(-c\log^2 n)$, $\forall\, c < c(\varepsilon)$.

(b) *If* $k \geq (2+a)\log^2 n$, $a > 0$, *then a stable matching exists with probability* $1 - O(n^{-c})$, $\forall\, c < c(a)$.

[*See* (6.2) *and* (6.4) *for* $c(\varepsilon)$ *and* $c(a)$.]

Thus, in the random graphs terminology (see Erdös and Rényi [6], Bollobás [1] and Palmer [21]), $k(n) = \log^2 n$ is a threshold function for existence of a stable matching. In contrast, for a complete matching property in a random bipartite graph $G = (V, E)$, $V = (V_1, V_2)$, $|V_1| = |V_2| = n$, $\deg(v) = k$, $\forall\, v \in V_1$), the (sharp) threshold function is $\log n$ ([6], [1]).

APPENDIX

Let $\mathscr{L}_1^{(n)}, \dots, \mathscr{L}_n^{(n)}$ be the lengths of consecutive subintervals obtained by selecting uniformly at random $(n-1)$ points in the interval $[0,1]$.

PROPOSITION A.1. *If* $x_i \in [0,1]$, $1 \leq i \leq n$, *are such that* $\sum_{i=1}^n x_i \geq 1$, *then*

$$P\left(\mathscr{L}_1^{(n)} \leq x_1, \dots, \mathscr{L}_n^{(n)} \leq x_n\right) \leq \left(\sum_{i=1}^n x_i - 1\right)^{n-1}. \qquad \text{(A.1)}$$

PROOF. By induction on n. For $n = 2$, the inequality (A.1) holds as equality since

$$P\left(\mathscr{L}_1^{(2)} \le x_1, \mathscr{L}_2^{(2)} \le x_2\right) = \int_{1-x_1}^{x_2} dx = x_2 + x_1 - 1.$$

Suppose that (A.1) is true for $n = m$. Since the joint density of $\mathscr{L}_1^{(m+1)}, \dots, \mathscr{L}_{m+1}^{(m+1)}$ equals $m!$, we have

$$\begin{aligned} &P\left(\mathscr{L}_1^{(m+1)} \le x_1, \dots, \mathscr{L}_m^{(m+1)} \le x_m, \mathscr{L}_{m+1}^{(m+1)} \le x_{m+1}\right) \\ \text{(A.2)} \quad &= m! \int_{\{y_i \le x_i:\ 1 \le i \le m+1\}} dy_1 \cdots dy_m \qquad \left(y_{m+1} = 1 - \sum_{i=1}^{m} y_i\right) \\ &= m \int_x^{x_1} P\left(\mathscr{L}_1^{(m)} \le x_2, \dots, \mathscr{L}_{m-1}^{(m)} \le x_m, \mathscr{L}_m^{(m)} \le x_{m+1} + y_1\right) dy_1, \end{aligned}$$

where

$$x = \max\left(0, 1 - \sum_{i=2}^{m+1} x_i\right).$$

By the induction hypothesis, the quantity in the last line of (A.2) is at most

$$m \int_x^{x_1} \left(\sum_{i=2}^{m+1} x_i + y_1 - 1\right)^{m-1} dx \le \left(\sum_{i=1}^{m+1} x_i - 1\right)^m. \qquad \square$$

COROLLARY. *For every* $x \ge n^{-1}$,

$$P\left(\max_{1 \le i \le n} \mathscr{L}_i^{(n)} \le x\right) \le (nx - 1)^{n-1}.$$

PROPOSITION A.2. *For* $\varepsilon \in (0, 1)$,

$$P\left(\max_{1 \le i \le n} \mathscr{L}_i^{(n)} \le (1 - \varepsilon) \frac{\log n}{n}\right) = O(\exp(-n^\rho)), \qquad \forall \rho \in (0, \varepsilon).$$

The proof is very much similar to the second part of the proof of Lemma 2 in Pittel [22]. A key point is the well-known fact that

$$\left\{\mathscr{L}_i^{(n)}: 1 \le i \le n\right\} \equiv_{\mathscr{D}} \left\{\frac{W_i}{\sum_{k=1}^n W_k}: 1 \le i \le n\right\},$$

where $W_1, \dots, W_n$ are i.i.d. random variables with the distribution function $1 - e^{-x}$.

NOTE. Devroye [3] (see also [4]) was able to find sharp asymptotic bounds of the preceding probability for smaller deviations, when $\varepsilon \to 0$ but $\varepsilon \log n \to \infty$, in a more general case of the kth longest subinterval. Devroye used these bounds to investigate, rather thoroughly, the almost-sure behavior of this random variable for $n \to \infty$.

Acknowledgments. I am very much indebted to Don Knuth for an invitation to spend the year of 1988 in Stanford. He introduced me to probabilistic aspects of the stable marriage problem and pointed out, in particular, that his integral formula (3.1) might be a key ingredient of asymptotic analysis. Many discussions of proposal algorithms with Don Knuth and Rajeev Motwani were tremendously helpful to me, and I greatly enjoyed our teamwork on the subject of stable husbands. Two years later these problems continue to fascinate me as much as they did when I learned of them from Don.

My thanks to Doug Critchlow, Alan Frieze, Volodya Lifschitz, Peter March, Louis Sucheston, Joe Verducci and Ricky Vohra for their friendly interest in the present paper. My love and gratitude to Irina, my wife, whose patience and sense of humor helped me so much during many months of work on the problems of stable marriages.

It is my pleasure to thank the referees and an Associate Editor who meticulously read the paper and came up with many insightful comments and helpful suggestions.

Note added in proof. Recently while studying a *nonbipartite* stable matching problem [B. Pittel (1992), On a random instance of the "stable roommates" problem, I, II, unpublished] we have found a way to close the gap in these bounds. It turns out that whp $\mathscr{Q}(\mathscr{M}_1)$, $\mathscr{R}(\mathscr{M}_2)$ are asymptotic to $\log^2 n$, and $\mathscr{Q}(\mathscr{M}_i)$, $\mathscr{R}(\mathscr{M}_i)$ are asymptotic to $n^{1/2} \log n$, $i = 3, 4$. ($\mathscr{M}_4$ is the minimum-regret stable matching.) The proof shows also that the total spouse ranks of $\mathscr{M}_3$ and $\mathscr{M}_4$ are, indeed, close to each other, with superpolynomially high probability.

REFERENCES

[1] BOLLOBÁS, B. (1985). *Random Graphs*. Academic, London.

[2] CHERNOFF, H. (1952). A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Ann. Math. Statist.* **23** 493–507.

[3] DEVROYE, L. (1981). Laws of the iterated logarithm for order statistics of uniform spacings. *Ann. Probab.* **9** 860–867.

[4] DEVROYE, L. (1982). A log log law for maximal uniform spacings. *Ann. Probab.* **10** 863–868.

[5] DURRETT, R. (1991). *Probability, Theory and Examples*. Brooks/Cole, Pacific Grove, Calif.

[6] ERDÖS, P. and RÉNYI, A. (1960). On the evolution of random graphs. *Magyar Tud. Akad. Mat. Kut. Int. Közl.* **5** 17–61.

[7] FEDER, T. (1989). A new fixed point approach for stable networks and stable marriages. In *Proc. 21st ACM Symp. on Theory of Computing* 513–522. Assoc. Comput. Mach., New York.

[8] FEDER, T. (1991). Network flow and 2-satisfiability. *Algorithmica*. To appear.

[9] FELLER, W. (1973). *An Introduction to Probability Theory and Its Applications* **2**. Wiley, New York.

[10] GALE, D. and SHAPLEY, L. S. (1962). College admissions and the stability of marriage. *Amer. Math. Monthly* **69** 9–15.

[11] GUSFIELD, D. (1987). Three fast algorithms for four problems in stable marriage. *SIAM J. Comput.* **16** 111–128.

[12] GUSFIELD, D. and IRVING, R. W. (1989). *The Stable Marriage Problem (Structure and Algorithms)*. MIT Press.

[13] IRVING, R. W. (1985). An efficient algorithm for the stable room-mates problem. *J. Algorithms* **6** 577–595.

[14] IRVING, R. W. and LEATHER, P. (1986). The complexity of counting stable marriages. *SIAM J. Comput.* **15** 655–667.

[15] IRVING, R. W., LEATHER, P. and GUSFIELD, D. (1987). An efficient algorithm for the "optimal" stable marriage. *J. Assoc. Comput. Mach.* **34** 532–543.

[16] ITOGA, S. Y. (1978). The upper bound for the stable marriage problem. *J. Oper. Res. Soc.* **29** 811–814.

[17] KNUTH, D. E. (1976). *Marriages Stables et Leurs Relations avec d'Autres Problèmes Combinatoires*. Les Presses de l'Université de Montréal.

[18] KNUTH, D. E. (1988). Personal communication.

[19] KNUTH, D. E., MOTWANI, R. and PITTEL, B. (1990). Stable husbands. *Random Structures and Algorithms* **1** 1–14.

[20] MCVITIE, D. G. and WILSON, L. B. (1971). The stable marriage problem. *Comm. ACM* **14** 486–490.

[21] PALMER, E. M. (1985). *Graphical Evolution*. Wiley, New York.

[22] PITTEL, B. (1989). The average number of stable matchings. *SIAM. J. Discrete Math.* **2** 530–549.

[23] WILSON, L. B. (1972). An analysis of the stable assignment problem. *BIT* **12** 569–575.

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210