The Annals of Probability
1993, Vol. 21 No. 3, 1441–1477

THE "STABLE ROOMMATES" PROBLEM WITH RANDOM PREFERENCES[1]

BY BORIS PITTEL

Ohio State University

In a set of even cardinality n, each member ranks all the others in order of preference. A stable matching is a partition of the set into $n/2$ pairs, with the property that no two unpaired members both prefer each other to their partners under matching. It is known that for some problem instances no stable matching exists. What if an instance of the ranking system is chosen uniformly at random? We show that the mean and the variance of the total number of stable matchings for the random problem instance are asymptotic to $e^{1/2}$ and $(\pi n/4e)^{1/2}$, respectively. Consequently, Prob (a stable matching exists) $\gtrsim (4e^3/\pi n)^{1/2}$. We also prove that, given the last event, in every stable matching the sum of the ranks of all members (as rank ordered by their partners) is asymptotic to $n^{3/2}$, and the largest rank of a partner is of order $n^{1/2} \log n$, with superpolynomially high conditional probability. In other words, stable partners are very likely to be relatively close to the tops of each other's preference lists.

1. Introduction, history and main results. Consider a set I of even cardinality n, and assume that for each member $a \in I$ there is given a subset $I(a)$ of "acceptable" partners ranked (without ties) according to a's individual preferences. A stable matching problem is to find a partition of I into $n/2$ admissible pairs with the property that no two unpaired members both prefer each other to their partners under matching. In 1962, Gale and Shapley [8] introduced and studied a "stable marriages" problem, which is a bipartite version of the stable matching problem, that is $I = I_1 + I_2$, $|I_1| = |I_2| = n/2$ and $I(a) = I_2$ or I_1 dependent upon whether $a \in I_1$ or $a \in I_2$. The sets I_1, I_2 can be interpreted as the set of men and the set of women, and a matching between I_1 and I_2 as a set of marriages (a marriage assignment). Gale and Shapley described a finite-step algorithm; it consists of rounds of proposals of currently free men, each to his next best woman, with every collision of suitors resolved in favor of the better suitor and all the rejected men having to propose in the next round, (see [8] for the complete description). They proved that a final round results in a stable matching, thus establishing that at least one such matching (marriage assignment) always exists. In 1971, McVitie and Wilson [20] suggested an alternative proposal algorithm in which, unlike Gale–Shapley's algorithm, the proposals by men are made one at a time.

Received January 1991; revised June 1992.
[1]Research partially supported by NSF Grant DMS-90-02347.
AMS 1991 *subject classifications*. Primary 60C05, 60F99; secondary 05A05, 05C70, 05C80, 41A60, 41A63.
Key words and phrases. stable matching, random preferences, asymptotic mean, variance, ranks, order, conditional, probability.

Remarkably, this algorithm yields the same stable matching, and it requires in the course of its work the same set of proposals as the algorithm in [8].

For every n, however, there are problem instances with only one stable matching. (Can the reader think of an example?) As for a nonbipartite version, the authors of [8] found a simple instance of a "stable roommates" problem $[I(x) = I \setminus \{x\};\ x \in I]$, for which no stable matching exists. Years later, it was found that, for both the stable marriage problem and the stable roommates problem, the maximal number of stable matchings grows (at least) exponentially fast with n (Knuth [16, 17], Irving and Leather [12] and Gusfield and Irving [10]).

Since the number of stable matchings varies so greatly from one instance to another, it is natural to ask how large this number is "typically," that is, in the case when rankings are chosen independently and uniformly from possible rankings of acceptable partners. For the stable marriage problem, this was one of the questions posed by Knuth in his book [16]; he suggested that it might be possible to estimate the average number of stable matchings via asymptotic study of his multidimensional integral-type formula for the probability that a given matching is stable. We performed this analysis in [21]; it turned out that $E(S_n)$ (the expected value of S_n, the total number of stable marriages) is asymptotic to $n \log n/2e$. Besides, we established that with high probability (whp) the two extreme stable matchings—male optimal and female optimal—dramatically differ from each other, whence $S_n \geq 2$ whp.

Soon after, Knuth, Motwani and Pittel [18] proved that whp $S_n \geq (1/2 - \varepsilon)\log n$, $\forall\ \varepsilon > 0$, that is, whp S_n is *unbounded*. The proof was based on a careful probabilistic analysis of a version of the McVitie–Wilson algorithm [20], which determines *sequentially* all stable partners of a particular woman; it was shown that whp t_n, the total number of those men is at least $(1/2 - \varepsilon)\log n$. In [22], we suggested another way to analyze this algorithm. It turned out that t_n is asymptotically Gaussian with mean and variance $\sim \log n$. Thus, whp $S_n \geq t_n \geq (1 - \varepsilon)\log n$, $\forall\ \varepsilon > 0$. Subsequently, we were able to narrow somewhat this glaring gap between the whp lower bound of S_n and its expectation $E(S_n) \sim n \log n/2e$. Using the results of Irving and Leather on the structure of stable marriages [13], and extending the techniques of our paper [21], we proved in [22] that whp $S_n \gtrsim (n/2 \log n)^{1/2}$.

Regardless of the sharpness of the last bound, one thing is already clear: Typically, the stable marriage problem has plenty of solutions, $n^{1/2+o(1)}$ at least. What about the random instance of the stable roommates problem?

In 1985, Irving [11] solved a difficult problem proposed by Knuth [16] in 1976. Irving found a polynomial-time in the worst case $[O(n^2)]$ algorithm that determines, for any instance of the stable roommates problem, whether a stable matching exists, and if so, finds such a matching. The algorithm has two phases, a proposal phase (not unlike McVitie–Wilson's algorithm), and a second phase based on an interesting notion of a stable table and rotations exposed in it. Having made experimental runs of the algorithm, Irving concluded that the proportion of problem instances with a stable matching $[P(S_n \geq 1)]$ appears to be a decreasing function of n. However, the decrease rate seems to be rather slow, so that it would be difficult to choose between two competing conjectures, $\lim_{n\to\infty} P(S_n \geq 1) = 0$ or $\lim_{n\to\infty} P(S_n \geq 1) > 0$.

In this article, (Section 3), we show that $\lim_{n\to\infty} E(S_n) = e^{1/2}$, so that the expected number of stable matchings is bounded, in a striking contrast with the case of stable marriages [21]. [Rob Irving has informed me that his student's results of simulations obtained several years ago indicated that $E(S_n)$ had to be a bounded function of n.] Thus, S_n is bounded in probability as $n \geq \infty$. But what about the value of $\lim_{n\to\infty} P(S_n \geq 1)$? In an attempt to answer this question, we decided to study the asymptotic behavior of the second order moment $E(S_n^2)$. Had that turned out to be bounded also, we would have been able to conclude [by means of Cauchy's inequality $P(S_n \geq 1) \geq E^2(S_n)/E(S_n^2)$] that $P(S_n \geq 1)$ is bounded away form 0. However, this moment happens to be unbounded; namely $E(S_n^2) \sim (\pi n/4e)^{1/2}$, as $n \to \infty$. So, we can assert only that

$$P(S_n \geq 1) \gtrsim \frac{2e^{3/2}}{\pi^{1/2} n^{1/2}}, \qquad n \to \infty. \tag{1.1}$$

Thus, it remains unclear whether $P(S_n \geq 1)$ converges to zero. In the case it does, the rate of convergence is quite slow. At this moment, we have a feeling that $P(S_n \geq 1) \to 0$, which is conjectured by Gusfield and Irving in their recently published book [10]. (See Note added in proof at the end of the text.)

We also demonstrate (Section 4) that only with superpolynomially low probability may the random problem instance have a stable matching with the sum of the ranks of all members (as ranked by their partners) being outside $[(1-\varepsilon)n^{3/2}, (1+\varepsilon)n^{3/2}]$, or the largest rank of a member being outside $[(1-\varepsilon)n^{1/2}\log n, (1+\varepsilon)n^{1/2}\log n]$, $\forall\, \varepsilon > 0$. So, by (1.1), given that a stable matching exists, with superpolynomially high *conditional* probability for every stable matching the sum of the ranks of all members is close to $n^{3/2}$, while the largest rank of a member is of order $n^{1/2}\log n$, precisely. Loosely speaking, all the stable matchings (whenever they exist) are very likely to be well balanced. This provides another illustration of how sharply the stable roommates problem differs from the stable marriages problem. We had earlier proved [22] a very likely existence of a large variety of stable marriage assignments, with the total spouses' rank ranging from $(n^3/2)^{1/2}$ to $n^2/4\log n$, and the largest rank of a spouse ranging from $(n/2)^{1/2}\log n$ to $n/2$.

Concluding, we should note that there is a conceptual (and mathematical) similarity between the problems on likely structure of stable matchings and a class of problems in population genetics, concerning likely size and shape of stable polymorphisms (under an assumption that the various genotypic fitnesses are independently random); see Karlin [13] and Kingman [14, 15].

2. Integral formulas. There are $[(n-1)!]^n$ instances of the ranking system in the case of the stable roommates problem, since every member $a \in I$ can order the remaining $(n-1)$ members in $(n-1)!$ ways. The random instance, chosen according to the uniform distribution, can be generated as follows. Introduce an $n \times n$ array of the independent random variables $X_{ij}(1 \le i \ne j \le n)$, each uniformly distributed on $[0, 1]$. Postulate that each member i ranks the members of $I \setminus \{i\}$ in the increasing order of the variables $X_{ij}(j \in I \setminus \{i\})$. Obviously, such an ordering is *uniform* for every i, and the orderings by different members are *independent*.

Assuming that $I = \{1, 2, \dots, n\}$, introduce a standard matching $M_0 = \{(i, i + n/2)\colon 1 \le i \le n/2\}$. Denote the probability that M_0 is stable by P_n, and the probability that M_0 is stable and the sum of the ranks of all members in M_0 equals k by P_{nk}. By symmetry, the values of these probabilities for any other fixed matching are P_n and P_{nk}, too. The random scheme we introduced above allows us to represent P_n and P_{nk} as the values of certain multidimensional integrals, not unlike those in Knuth [16], and Pittel [21] for the stable marriages (cf. Kingman [14, 15]).

LEMMA 1. *Let C denote the n-dimensional unit cube, that is, $C = \{x = (x_1, \dots, x_n)\colon 0 \le x_i \le 1,\ 1 \le i \le n\}$. Then*

$$P_n = \int_C \prod_{\{i, j\} \in M_0^c} (1 - x_i x_j)\, dx, \qquad dx = dx_1 \dots dx_n, \tag{2.1}$$

and, for $k \ge n$,

$$P_{nk} = \int_C [z^{k-n}] \prod_{\{i, j\} \in M_0^c} \left(\bar{x}_i \bar{x}_j + z x_i \bar{x}_j + z \bar{x}_i x_j\right) dx \tag{2.2}$$

($\bar{x}_\alpha = 1 - x_\alpha$). *Both products are taken over all* unordered *pairs of distinct members of I that do not form a match in M_0. The integrand in* (2.2) *is the coefficient of z^{k-n} in the polynomial expansion of the product.*

PROOF. For clarity, we give separate proofs of the relations (2.1) and (2.2), even though the former is a consequence of the latter. (Does the reader see why?)

(a) By the definition of stability,

$$\{M_0 \text{ is stable}\} = \bigcap_{\{i, j\} \in M_0^c} A_{ij}^c, \tag{2.3}$$

where

$$A_{ij} = \{X_{ij} < X_{i, i+n/2};\ X_{ji} < X_{j, j+n/2}\}; \tag{2.4}$$

(both $i + n/2$ and $j + n/2$ are taken modulo n). Now, all $X_{\alpha\beta}$ are independent (and uniform on $[0, 1]$). So, by Fubini's theorem, conditioned on $X_{\alpha, \alpha + n/2} = x_\alpha (1 \le a \le n)$, the events A_{ij}, $\{i, j\} \in M_0^c$, are independent, and

$$P(A_{ij} | \cdot) = x_i x_j.$$

$[P(\ | \cdot)$ here, and $E(\ | \cdot)$ below denote the conditional probability and the conditional expectation.] Therefore,

$$P(M_0 \text{ is stable} \mid \cdot) = \prod_{\{i,j\}\in M_0^c} (1 - x_i x_j)$$

and (2.1) follows.

(b) Let R denote the sum of the ranks of all members in M_0, and $\chi(\mathscr{A})$, [or $\chi_{\mathscr{A}}(\cdot)$] the set indicator of an event $\mathscr{A}$. Then

$$R = n + \sum_{\{i,j\}\in M_0^c} \left[\chi(X_{ij} < X_{i,i+n/2}) + \chi(X_{ji} < X_{j,j+n/2})\right], \tag{2.5}$$

and

$$P_{nk} = [z^k]E\left[z^R \chi(M_0 \text{ is stable})\right]. \tag{2.6}$$

Further, by (2.3)–(2.5) and Fubini's theorem, again conditioning on $X_{\alpha,\alpha+n/2} = x_\alpha$,

$$\begin{aligned} &E\left[z^{R-n}\chi(M_0 \text{ is stable}) \mid \cdot\right] \\ &= E\left[\left(\prod_{\{i,j\}\in M_0^c} \chi(A_{ij}^c) z^{\chi(X_{ij} < X_{i,i+n/2}) + \chi(X_{ji} < X_{j,j+n/2})}\right) \Big| \cdot \right] \\ &= \prod_{\{i,j\}\in M_0^c} E\left[\chi\left(\{X_{ij} < x_i;\ X_{ji} < x_j\}^c\right) z^{\chi(X_{ij} < x_i) + \chi(X_{ji} < x_j)}\right] \\ &= \prod_{\{i,j\}\in M_0^c} \left(\bar{x}_i \bar{x}_j + z x_i \bar{x}_j + z \bar{x}_i x_j\right). \end{aligned}$$

This and (2.6) imply (2.2). □

Continuing, suppose that we have a pair of distinct matchings, M_1 and M_2. How do we compute the probability $P(M_1, M_2)$ that both matchings are stable? Together, M_1 and M_2 determine a graph $G(M_1, M_2) = (I, E)$, $E = M_1 \cup M_2$, that is, the edge set is formed by pairs $\{i, j\}$ from M_1 or M_2. A component of G is either an edge, and the set of all such edges is $M_1 \cap M_2$, or a circuit of even length greater than or equal to 4 in which the edges from M_1 and M_2 alternate. The edge set for all these circuits is $M_1 \Delta M_2 = (M_1 \setminus M_2) \cup (M_2 \setminus M_1)$. We call the circuits alternating, by analogy with alternating paths in matching theory (see, e.g., Swamy and Thulasiraman [26]).

LEMMA 2. *Given a graph $G = G(M_1, M_2)$, let D be the set of all $(x, y) \in C \times C$ such that* (a) $x_\alpha = y_\alpha$ *for every noncyclic vertex α, and* (b) *for every circuit $\{i_1, i_2, \dots, i_m\}$ of the graph G either*

$$x_{i_1} > y_{i_1}, x_{i_2} < y_{i_2}, \dots, x_{i_m} < y_{i_m},$$

or

$$x_{i_1} < y_{i_1}, x_{i_2} > y_{i_2}, \dots, x_{i_m} > y_{i_m}$$

hold. (Recall that each m is even.) Then, denoting $\min(a, b)$ *by $a \wedge b$, and the* *matching function for* M_α *by* $m_\alpha(\cdot)$, $\alpha = 1, 2$,

$$P(M_1, M_2) = \int_D \prod_{\{i,j\} \in M_1^c \cap M_2^c} \left[1 - x_i x_j - y_i y_j + (x_i \wedge y_i)(x_j \wedge y_j)\right] dx\,dy,$$

$$dx = \prod_{\alpha=1}^{n} dx_\alpha, \qquad dy = \prod_{\alpha:\, m_1(\alpha) \neq m_2(\alpha)} dy_\alpha.$$

PROOF. By the definition of stability, we have

$$\{M_1, M_2 \text{ are both stable}\} = \bigcap_{\{i,j\} \in (M_1 \cap M_2)^c} B_{ij}^c.$$

Here, if $\{i, j\} \in M_1^c \cap M_2^c$,

$$B_{ij} = \{X_{ij} < X_{i, m_1(i)};\ X_{ji} < X_{j, m_1(j)}\} \cup \{X_{ij} < X_{i, m_2(i)};\ X_{ji} < X_{j, m_2(j)}\};$$

if $\{i, j\} \in M_1^c \cap M_2$, then $j = m_2(i)$ and $i = m_2(j)$ and

$$B_{ij} = \{X_{i, m_2(i)} < X_{i, m_1(i)};\ X_{j, m_2(j)} < X_{j, m_1(j)}\};$$

likewise for $\{i, j\} \in M_1 \cap M_2^c$,

$$B_{ij} = \{X_{i, m_1(i)} < X_{i, m_2(i)};\ X_{j, m_1(j)} < X_{j, m_2(j)}\}.$$

Conditioned on the values $X_{\alpha, m_1(\alpha)} = x_\alpha$, $X_{\alpha, m_2(\alpha)} = y_\alpha$, $1 \le \alpha \le n$, the events B_{ij} (whence B_{ij}^c) are independent, and

$$P\left(B_{ij}^c \mid \cdot\right) = \begin{cases} 1 - x_i x_j - y_i y_j + (x_i \wedge y_i)(x_j \wedge y_j), & \{i, j\} \in M_1^c \cap M_2^c, \\ \chi(\{y_i \ge x_i, \text{ or } y_j \ge x_j\}), & \{i, j\} \in M_1^c \cap M_2, \\ \chi(\{x_i \ge y_i, \text{ or } x_j \ge y_j\}), & \{i, j\} \in M_1 \cap M_2^c. \end{cases}$$

Therefore,

$$P(M_1, M_2 \text{ are both stable} \mid \cdot) = \prod_{\{i,j\} \in M_1^c \cap M_2^c} \left[1 - x_i x_j - y_i y_j + (x_i \wedge y_i)(x_j \wedge y_j)\right],$$

under the condition "$\{i, j\} \in M_1^c \cap M_2 \Rightarrow y_i \ge x_i$, or $y_j \ge x_j$ and $\{i, j\} \in M_1 \cap M_2^c \Rightarrow x_i \ge y_i$, or $x_j \ge y_j$." If the condition does not hold, the conditional probability is zero. Since the edges from $M_1 \Delta M_2$ form the disjoint alternating cycles in the graph G, the condition means that for every such cycle $\{i_1, i_2, \dots, i_m\}$ [with $(i_1, i_2) \in M_1$, $(i_2, i_3) \in M_2, \dots, (i_m, i_1) \in M_2$, say]

$$\begin{aligned} y_{i_1} \le x_{i_1} &\quad \text{or} \quad y_{i_2} \le x_{i_2}, \\ y_{i_2} \ge x_{i_2} &\quad \text{or} \quad y_{i_3} \ge x_{i_3}, \\ &\ \ \vdots \\ y_{i_{m-1}} \le x_{i_{m-1}} &\quad \text{or} \quad y_{i_m} \le x_{i_m}, \\ y_{i_m} \ge x_{i_m} &\quad \text{or} \quad y_{i_1} \ge x_{i_1}. \end{aligned} \tag{2.7}$$

Since $m_1(i_j) \neq m_2(i_j)$ for $1 \le j \le m$ on an alternating circuit, $y_{i_j} = X_{i_j, m_2(i_j)} \neq X_{i_j, m_1(i_j)} = x_{i_j}$ almost surely for these j. But if $y_{i_j} \neq x_{i_j}$, $1 \le j \le m$, then the conditions (2.7) are met if either $x_{i_1} > y_{i_1}, x_{i_2} < y_{i_2}, \dots, x_{i_m} < y_{i_m}$, or $x_{i_1} < y_{i_1}, x_{i_2} > y_{i_2}, \dots, x_{i_m} > y_{i_m}$. (The inequalities alternate.) This proves the lemma. □

3. Two moments of the number of stable matchings. Let S_n stand for the total number of stable matchings. We use the results of Section 2 in order to obtain asymptotics of the first two moments, $E(S_n)$ and $E(S_n^2)$.

THEOREM 1. $\lim_{n\to\infty} E(S_n) = e^{1/2}$; *so, in particular,* S_n *is bounded in probability, as* $n \to \infty$.

PROOF. Since the total number of matchings is $1 \cdot 3 \cdot 5 \cdot \dots \cdot (n-1)$ ($=_{\text{def}} (n-1)!!$), we have

$$E(S_n) = (n-1)!!P_n.$$

Here $P_n = P(M_0$ is stable) is given by (see Lemma 1)

$$P_n = \int_C \Pi(x)\,dx, \tag{3.1}$$

where

$$\Pi(x) = \prod_{\{i,j\}\in M_0^c} (1 - x_i x_j).$$

To estimate P_n, and elsewhere, we will frequently use the following bound (cf. Pittel [22]).

LEMMA 3. *Define* $s = \sum_{i=1}^n x_i$, *and* $v = \{v_i = x_i/s;\ 1 \le i \le n\}$, *so that* $\sum_{i=1}^n v_i = 1$. *Let* $L^{(n)} = \{L_i^{(n)};\ 1 \le i \le n\}$ *be the set of lengths of the consecutive subintervals of* $[0,1]$ *obtained by selecting, independently and uniformly at random,* $(n-1)$ *points in* $[0,1]$. *Then for nonnegative (measurable) functions* $f(s), g(v)$,

$$\begin{aligned}\int_C f(s)g(v)\,dx &= \int_0^n f(s)\frac{s^{n-1}}{(n-1)!}E\Big(g(L^{(n)})\chi\Big(\max_i L_i^{(n)} \le s^{-1}\Big)\Big)\,ds \\ &\le E\big(g(L^{(n)})\big)\int_0^n f(s)\frac{s^{n-1}}{(n-1)!}\,ds,\end{aligned} \tag{3.2}$$

with equality if $sv \in C$ *whenever* $f(s) > 0$ *and* $g(v) > 0$.

More generally, if $\{I_j;\ 1 \le j \le m\}$ is a partition of $\{1,\dots,n\}$, then introducing $s_j = \sum_{i\in I_j} x_i$, $v^{(j)} = \{x_i/s_j;\ i \in I_j\}$, $n_j = |I_j|$,

$$\begin{aligned}\int_C f(s_1,\dots,s_m)\left(\prod_{j=1}^m g_j(v^{(j)})\right)dx\\ &= \overbrace{\int\cdots\int}^{k}_{0\le s_j\le n_j} f(s_1,\dots,s_m)\\ &\quad\times\left[\prod_{j=1}^m \frac{s_j^{n_j-1}}{(n_j-1)!}E\left(g(L^{(n_j)})\chi\left(\max_{1\le i\le n_j} L_i^{(n_j)} \le s_j^{-1}\right)\right)\right]\\ &\quad\times ds_1\cdots ds_m.\end{aligned}\tag{3.3}$$

The proof is short. On the left-hand side in (3.2), we switch from x to s, $v_1,\dots,v_{n-1}$ (so that $v_i \le s^{-1}$, $1 \le i \le n$), notice that the Jacobian of the inverse transform equals s^{n-1}, and then recall that $\rho(v_1,\dots,v_{n-1})$, the density of $(L_1^{(n)},\dots,L_{n-1}^{(n)})$, is given by

$$\rho = \begin{cases}(n-1)!, & \text{if } v\ge 0,\ \sum_{i=1}^{n-1} v_i \le 1,\\ 0, & \text{otherwise,}\end{cases}$$

(Breiman [2]).

The relation (3.2) clearly implies (3.3) for $f = \prod_j f_j(s_j)$. A usual extension argument yields then (3.3) for an arbitrary f. □

The bounds (3.2) and (3.3) are particularly effective in combination with a well-known fact; namely, that

$$L^{(n)} \equiv_{\mathscr{D}} \left\{\frac{w_i}{\sum_{j=1}^n w_j};\ 1 \le i \le n\right\},\tag{3.4}$$

where $w_1,\dots,w_n$ are exponential (with parameter 1, say), independent random variables, (Breiman [2] and Rényi [24]).

For clarity, we break the remaining argument into steps.

STEP 1. Since $1-\gamma \le e^{-\gamma}$, we have

$$\begin{aligned}\Pi(x) &\le \exp\left(-\sum_{\{i,j\}\in M_0^c} x_i x_j\right)\\ &= \exp\left(-\tfrac{1}{2}s^2 + \tfrac{1}{2}\sum_{i=1}^n x_i^2 + \sum_{i=1}^{n/2} x_i x_{i+n/2}\right)\\ &\le \exp\left(-\tfrac{1}{2}s^2 + \sum_{i=1}^n x_i^2\right) \le \exp\left(-\tfrac{1}{2}s^2 + s\right).\end{aligned}\tag{3.5}$$

Besides, by Stirling's formula,

$$(3.6)\qquad (n-1)!! = \frac{n!}{2^{n/2}\left(\frac{n}{2}\right)!} = O\left(\left(\frac{n}{e}\right)^{n/2}\right).$$

Define

$$(3.7)\qquad C_1 = \{x \in C:\ s \le n^{1/2}\log n\};$$

then

$$(n-1)!!\int_C \Pi(x)\,dx = (n-1)!!\int_{C_1}\Pi(x)\,dx + O\left(\exp\left(-\tfrac{1}{3}n\log^2 n\right)\right),$$

because, by (3.5)–(3.7),

$$(n-1)!!\int_{C-C_1}\Pi(x)\,dx = O\left(\exp\left(-\tfrac{1}{2}n\log^2 n + n^{1/2}\log n + \tfrac{1}{2}n\log\frac{n}{e}\right)\right).$$

STEP 2. Introduce $t_i(v)$, $i = 1, 2, 3$, as follows:

$$(3.8)\qquad t_1 = \sum_{i=1}^{n} v_i^2, \qquad t_2 = \sum_{i=1}^{n/2} v_i v_{i+n/2}, \qquad t_3 = \sum_{i=1}^{n} v_i^4.$$

Define C_2 as a subset of C_1 such that

$$(3.9)\qquad \frac{2}{n}(1-\varepsilon_n) \le t_1 \le \frac{2}{n}(1+\varepsilon_n),$$

$$(3.10)\qquad \frac{1}{2n}(1-\varepsilon_n) \le t_2 \le \frac{1}{2n}(1+\varepsilon_n)$$

and

$$(3.11)\qquad t_3 \le 25n^{-3},$$

where

$$\varepsilon_n = n^{-1/4}\log^2 n.$$

Let us agree to read "$\neg$(3.9)," say, as the negation of the condition (3.9), that is, "$t_1 < (2/n)(1-\varepsilon_n)$ or $t_1 > (2/n)(1+\varepsilon_n)$." Denote by R_1, R_2 and R_3 the contributions to the value of $(n-1)!!\int_{C_1}\Pi\,dx$ made by $C_1 \cap \neg$(3.9), $C_1 \cap \neg$(3.10) and $C_1 \cap \neg$(3.11), respectively. Obviously,

$$(n-1)!!\int_{C_1}\Pi(x)\,dx = (n-1)!!\int_{C_2}\Pi(x)\,dx + O(R_1 + R_2 + R_3).$$

Let us estimate R_1. Using Lemma 3 with $g(v) = \chi_{\neg(3.9)}(v)$, and also (3.5), (3.7) and (3.8), we have

$$\begin{aligned} R_1 &\le (n-1)!!\exp(n^{1/2}\log n)\left(\int_0^\infty e^{-s^2/2}\frac{s^{n-1}}{(n-1)!}\,ds\right)\\ &\quad\times P\left(\left|\frac{t_1(L^{(n)})}{2n^{-1}}-1\right|>\varepsilon_n\right) \qquad \left[\int_0^\infty e^{-s^2/2}s^{n-1}\,ds=(n-2)!!\right]\\ &= \exp(n^{1/2}\log n)P\left(\left|\frac{t_1(L^{(n)})}{2n^{-1}}-1\right|>\varepsilon_n\right), \end{aligned} \tag{3.12}$$

where [see (3.4)]

$$t_1(L^{(n)}) = \sum_{i=1}^n \left(L_i^{(n)}\right)^2 \equiv_{\mathscr{D}} \left(\sum_{i=1}^n w_i^2\right)\Bigg/\left(\sum_{j=1}^n w_j\right)^2. \tag{3.13}$$

LEMMA 4. *If $Y_1, Y_2, \dots$ are i.i.d. random variables with $E(e^{zY}) < \infty$ for $|z|$ sufficiently small, then there exist $c > 0$ and $\delta_0 > 0$ such that*

$$P\left(\left|\frac{1}{m}\sum_{i=1}^m Y_i - E(Y)\right| \ge \delta\right) = O(\exp(-cm\delta^2)), \qquad \forall\, \delta \le \delta_0. \tag{3.14}$$

(The proof is standard, based on the Chernoff method [3].)

Now, $E(w_1) = 1$, $E(w_1^2) = 2$; so by this lemma,

$$\begin{aligned} P_1 &=_{\text{def}} P\left(\left|\frac{1}{n}\sum_{i=1}^n w_i - 1\right| \ge n^{-1/4}\log n\right)\\ &= O(\exp(-c_1 n^{1/2}\log^2 n)), \qquad c_1 > 0,\\ P_2 &=_{\text{def}} P\left(\left|\frac{1}{n}\sum_{i=1}^n w_i^2 - 2\right| \ge n^{-1/4}\log n\right)\\ &= O(\exp(-c_2 n^{1/2}\log^2 n)), \qquad c_2 > 0. \end{aligned} \tag{3.15}$$

Then, since $n^{-1/4}\log n \ll \varepsilon_n = n^{-1/4}\log^2 n$,

$$\begin{aligned} P\left(\left|\frac{t_1(L^{(n)})}{2n^{-1}}-1\right|>\varepsilon_n\right) &\le P_1 + P_2\\ &= O(\exp(-cn^{1/2}\log^2 n)), \quad c = c_1 \wedge c_2 > 0. \end{aligned} \tag{3.16}$$

Consequently, by (3.12),

$$R_1 = O(\exp(-c' n^{1/2}\log^2 n)), \qquad c' < c.$$

Similarly, since $E(w_1w_2) = E(w_1)E(w_2) = 1$,

$$R_2 = \exp(n^{1/2}\log n)P\left(\left|\frac{t_2(L^{(n)})}{2^{-1}n^{-1}} - 1\right| > \varepsilon_n\right)$$
$$= O\exp(-c''n^{1/2}\log^2 n)), \qquad c'' > 0,$$

and $[E(w_1^4) = 24]$,

$$R_3 = \exp(n^{1/2}\log n)P\left(\frac{t_3(L^{(n)})}{24n^{-3}} > \frac{25}{24}\right)$$
$$= O(\exp(-c'''n)), \qquad c''' > 0.$$

Therefore,

$$(n-1)!!\int_{C_1}\Pi(x)\,dx$$
$$= (n-1)!!\int_{C_2}\Pi(x)\,dx + O(\exp(-cn^{1/2}\log^2 n)), \qquad c > 0.$$

STEP 3. Using $1 - \gamma \le \exp(-\gamma - \gamma^2/2)$, we have on C_2:

$$\Pi(x) \le \exp\left[-\sum_{\{i,j\}\in M_0^c}\left(x_ix_j + \frac{1}{2}x_i^2x_j^2\right)\right]$$

$$(3.17)\qquad \begin{aligned} &= \exp\left[-\frac{1}{2}s^2 + \frac{1}{2}\sum_{i=1}^n x_i^2 + \sum_{i=1}^{n/2}x_ix_{i+n/2} - \frac{1}{4}\left(\sum_{i=1}^n x_i^2\right)^2 + O\left(\sum_{i=1}^n x_i^4\right)\right] \\ &= \exp\left[-\frac{1}{2}s^2(1 - t_1 - 2t_2) - \frac{1}{4}s^4t_1^2 + O\left(\frac{s^4}{n^3}\right)\right] \\ &= \exp\left[-\frac{1}{2}s^2\left(1 - \frac{3}{n}\right) - \frac{s^4}{n^2} + O(n^{-1/4}\log^6 n)\right]. \end{aligned}$$

Then, by Lemma 3,

$$(n-1)!!\int_{C_2}\Pi(x)\,dx \le [1 + O(n^{-1/4}\log^6 n)]$$
$$\times\frac{(n-1)!!}{(n-1)!}\int_0^\infty \exp\left[-\frac{1}{2}s^2\left(1 - \frac{3}{n}\right) - \frac{s^4}{n^2}\right]s^{n-1}\,ds.$$

The integrand of the one-dimensional integral attains its sharp maximum at a point $s = n^{1/2} + O(n^{-1/2})$, and a little reflection shows that the integral is asymptotic to

$$e^{-1}\int_0^\infty \exp\left[-\frac{1}{2}s^2\left(1-\frac{3}{n}\right)\right]s^{n-1}\,ds$$
$$= e^{-1}\left(1-\frac{3}{n}\right)^{-n/2}\int_0^\infty \exp\left(-\frac{s^2}{2}\right)s^{n-1}\,ds$$
$$= (n-2)!!\exp\left(\frac{1}{2}+O(n^{-1})\right).$$

So,

$$\limsup\,(n-1)!!\int_C \Pi(x)\,dx = \limsup(n-1)!!\int_{C_2}\Pi(x)\,dx \le e^{1/2}.$$

STEP 4. It remains to prove that

$$\liminf\,(n-1)!!\int_C \Pi(x)\,dx \ge e^{1/2}.$$

To this end, introduce C_3, the set of all $x \in C_2$ such that, in addition,

$$s \le 2n^{1/2}, \qquad \max_{1\le i\le n} v_i \le \frac{2\log n}{n}.$$

Notice that these new conditions imply that

$$\max_{1\le i\le n} x_i \le 4\frac{\log n}{n^{1/2}} = o(1), \qquad n\to\infty,$$

whence the condition "$\max_{1\le i\le n} x_i \le 1$" is met automatically. Using $1-x = \exp(-x - x^2/2 + O(x^3))$, $x\to 0$, we obtain [similarly to (3.17)] on C_3:

$$\Pi(x) = \exp\left[-\frac{1}{2}s^2(1-t_1-2t_2) - \frac{1}{4}s^4t_1^2 + O\left(\left(\sum_{i=1}^n x_i^3\right)^2\right)\right]$$
$$\ge \exp\left[-\frac{s^2}{2}\left(1-\frac{3}{n}\right) - \frac{s^4}{n^2} + O(n^{-1/4}\log^6 n) + O\left(\frac{\log^6 n}{n}\right)\right].$$

So, using Lemma 3 (the case of equality), within a factor $1 + O(n^{-1/4}\log^6 n)$,

$$(n-1)!!\int_{C_3}\Pi(x)\,dx$$
$$\ge (n-1)!!\,e^{-1}\int_{s\le 2n^{1/2}}\exp\left[-\frac{1}{2}s^2\left(1-\frac{3}{n}\right)\right]\frac{s^{n-1}}{(n-1)!}ds \tag{3.18}$$
$$\times P\left(\left\{\max_j L_j^{(n)} \le \frac{2\log n}{n}\right\}\cap B\right),$$

where

$$B = \left\{ \left| \frac{t_1(L^{(n)})}{2n^{-1}} - 1 \right| \le \varepsilon_n \right\} \cap \left\{ \left| \frac{t_2(L^{(n)})}{2^{-1}n^{-1}} - 1 \right| \le \varepsilon_n \right\} \cap \left\{ t_3(L^{(n)}) \le 25n^{-3} \right\}.$$

The integral is asymptotic to $e^{3/2}(n-2)!!/(n-1)!$. We also know that $P(B) \to 1$. Besides,

$$(3.19)\qquad \begin{aligned} P\left(\max_j L_j^{(n)} \le \frac{2\log n}{n} \right) &\ge 1 - nP\left(L_1^{(n)} \ge \frac{2\log n}{n} \right) \\ &= 1 - n\left(1 - \frac{2\log n}{n}\right)^{n-1} \\ &= 1 - O(n^{-1}). \end{aligned}$$

Therefore, the probability factor in (3.18) approaches 1, as $n \to \infty$.

NOTE. Actually, $\log n/n$ is the *sharp* probabilistic estimate of $\max L_j^{(n)}$ (Levy [19] and Darling [4]). Devroye [6, 7] analyzed in great detail the almost sure behavior of the kth largest subinterval $L_j^{(n)}$, for each fixed $k \ge 1$.

Consequently,

$$\begin{aligned} \liminf (n-1)!! \int_C \Pi(x)\,dx &\ge \liminf (n-1)!! \int_{C_3} \Pi(x)\,dx \\ &= e^{1/2}. \end{aligned}$$

The theorem that $\lim_{n\to\infty} E(S_n) = e^{1/2}$ is proven. □

Turn now to the second order moment. The computations are considerably more involved, but the above proof provides a starting insight into the structure of those stable matchings which are responsible for the dominant part of $E(S_n^2)$.

THEOREM 2. *$E(S_n^2) \sim (\pi n/4e)^{1/2}$, as $n \to \infty$.*

PROOF. To begin, we observe first that

$$E(S_n(S_n - 1)) = \sum_{M_1, M_2} P(M_1, M_2),$$

where, for two distinct matchings M_1, M_2, $P(M_1, M_2)$ is the probability that both M_1 and M_2 are stable. By Lemma 2,

$$(3.20)\qquad P(M_1, M_2) = \int_D \Pi(x,y)\,dx\,dy;$$

here

$$\Pi(x,y) = \prod_{\{i,j\}\in M_1^c \cap M_2^c} \left[1 - x_i x_j - y_i y_j + (x_i \wedge y_i)(x_j \wedge y_j)\right]$$

and $D = D(M_1, M_2) \subset C \times C$ is defined by the additional conditions

$$x_\alpha = y_\alpha, \qquad \alpha \notin \mathscr{C}, \tag{3.21}$$

$\mathscr{C} = \mathscr{C}(M_1, M_2)$ being the set of cyclic vertices of the graph $G = G(M_1, M_2)$; for every circuit $\{i_1, \dots, i_m\}$ of G, either

$$x_{i_1} > y_{i_1}, \qquad x_{i_2} < y_{i_2}, \dots, x_{i_m} < y_{i_m}, \tag{3.22}$$

or

$$x_{i_1} < y_{i_1}, \qquad x_{i_2} > y_{i_2}, \dots, x_{i_m} > y_{i_m}. \tag{3.23}$$

STEP 1. Introduce $s = \sum_i x_i$, $s_* = \sum_i y_i$. Since the (i, j)th factor in $\Pi(x, y)$ is at most $1 - (x_i x_j \vee y_i y_j)$ ($a \vee b =_{\text{def}} \max(a, b)$),

$$\begin{aligned}\Pi(x, y) &\le \exp\Bigl(-\tfrac{1}{2}s^2 + \tfrac{3}{2}\sum_i x_i^2\Bigr) \wedge \exp\Bigl(-\tfrac{1}{2}s_*^2 + \tfrac{3}{2}\sum_i y_i^2\Bigr)\\ &\le \exp\bigl(\tfrac{3}{2}n - \tfrac{1}{2}(s \vee s_*)^2\bigr).\end{aligned}$$

Also, the total number of ordered pairs of matchings (M_1, M_2) is $(n-1)!!((n-1)!! - 1) = O(n^n)$. Therefore,

$$\begin{aligned}\sum_{M_1, M_2} \int_{D \cap \{s \vee s_* \ge n^{1/2} \log n\}} \Pi(x, y)\, dx\, dy &= O\bigl(n^n \exp\bigl(\tfrac{3}{2}n - \tfrac{1}{2}n \log^2 n\bigr)\bigr)\\ &= O\bigl(\exp\bigl(-\tfrac{1}{3}n \log^2 n\bigr)\bigr).\end{aligned}$$

Thus

$$E(S_n(S_n - 1)) = \sum_{M_1, M_2} \int_{D_1} \Pi(x, y)\, dx\, dy + O\bigl(\exp\bigl(-\tfrac{1}{3}n \log^2 n\bigr)\bigr), \tag{3.24}$$

where $D_1 = D_1(M_1, M_2)$ is defined by

$$D_1 = D \cap \{s \vee s_* \le n^{1/2} \log n\}.$$

STEP 2. Let $2\nu = 2\nu(M_1, M_2)$ stand for $|\mathscr{C}(M_1, M_2)|$, the total length of all circuits in $G(M_1, M_2)$. Introduce $\nu_1 = n^{3/4} \log n$. We want to show that

$$\Sigma_n =_{\text{def}} \sum_{\nu \ge \nu_1} \int_{D_1} \Pi(x, y)\, dx\, dy = O\bigl(\exp(-cn^{1/2} \log n)\bigr), \qquad c > 0, \tag{3.25}$$

that is [see (3.24)],

$$E(S_n(S_n - 1)) = \sum_{\nu \le \nu_1} \int_{D_1} \Pi(x, y)\, dx\, dy + O\bigl(\exp(-cn^{1/2} \log n)\bigr). \tag{3.26}$$

[The sums in (3.25) and (3.26) are over all pairs (M_1, M_2) such that $\nu(M_1, M_2) \ge \nu_1$ and $\nu(M_1, M_2) \le \nu_1$, respectively.] The bound (3.25) implies that, with superpolynomially high probability, for every two stable matchings $\mathscr{M}^*$ and $\mathscr{M}^{**}$, at least $n - 2\nu_1 = n - 2n^{3/4} \log n$ members have the same partners in $\mathscr{M}^*$ and $\mathscr{M}^{**}$. Using this fact, we will be able later to replace $n - 2\nu_1$ by an essentially best bound $n - 2n^{1/2} \log n$.

To prove (3.25), we first write

$$\Pi(x,y) \le \exp\Bigg(-\sum_{\{i,j\}\in M_1^c\cap M_2^c}\big[x_ix_j + y_iy_j - (x_i\wedge y_i)(x_j\wedge y_j)\big]\Bigg)$$

$$\le \exp\Bigg(-\sum_{i\ne j}\big[x_ix_j + y_iy_j - (x_i\wedge y_i)(x_j\wedge y_j)\big] + \sum_{i=1}^n (x_i^2+y_i^2)\Bigg) \tag{3.27}$$

$$\le \exp\Bigg(-\tfrac{1}{2}s^2 - \tfrac{1}{2}s_*^2 + \tfrac{1}{2}s_{**}^2 + \tfrac{3}{2}\sum_{i=1}^n (x_i^2+y_i^2)\Bigg),$$

where $s_{**} = \sum_{i=1}^n (x_i \wedge y_i)$. Therefore, for $(x,y)\in D_1$,

$$\begin{aligned} &\Pi(x,y) \le \exp(3n^{1/2}\log n)\tilde{\Pi}(x,y),\\ &\tilde{\Pi}(x,y) =_{\text{def}} \exp\big(-\tfrac{1}{2}s^2 - \tfrac{1}{2}s_*^2 + \tfrac{1}{2}s_{**}^2\big). \end{aligned} \tag{3.28}$$

A crucial property of $\tilde{\Pi}$ is that, unlike Π, it depends only on the sums, s, s_* and s_{**}. This and the nature of the conditions (3.21)–(3.23) make it obvious that

$$\int_{D_1}\tilde{\Pi}(x,y)\,dx\,dy \le 2^{\mu(G)}\int_{D^*}\tilde{\Pi}(x,y)\Bigg(\prod_{\alpha=1}^n dx_\alpha \prod_{\beta=1}^{2\nu} dy_\beta\Bigg), \tag{3.29}$$

where $\mu(G)$ is the total number of circuits in G, $y_\alpha =_{\text{def}} x_\alpha$, $2\nu+1\le\alpha\le n$, and D^* is such that

$$\begin{aligned} y_\alpha \ge x_\alpha \ge 0, &\qquad 1\le\alpha\le\nu,\\ x_\alpha \ge y_\alpha \ge 0, &\qquad \nu+1\le\alpha\le 2\nu,\\ x_\alpha \ge 0, &\qquad 2\nu+1\le\alpha\le n. \end{aligned}$$

It is possible to evaluate the last integral precisely.

To achieve this, introduce $x' = \{x'_\alpha;\ 1\le\alpha\le n\}$, $y' = \{y'_\alpha\colon 1\le\alpha\le 2\nu\}$: $x'_\alpha = x_\alpha - y_\alpha$, $\nu+1\le\alpha\le 2\nu$, $x'_\alpha = x_\alpha$ otherwise; $y'_\alpha = y_\alpha - x_\alpha$, $1\le\alpha\le\nu$, $y'_\alpha = y_\alpha$, $\nu+1\le\alpha\le 2\nu$. In the new variables, D^* is the cone of all nonnegative (x',y'), and the Jacobian $\partial(x,y)/\partial(x',y')$ equals 1. Furthermore, a simple calculation shows that

$$-\tfrac{1}{2}s^2 - \tfrac{1}{2}s_*^2 + \tfrac{1}{2}s_{**}^2 = -\tfrac{1}{2}\Bigg(\sum_{\alpha=1}^n x'_\alpha + \sum_{\alpha=1}^{2\nu} y'_\alpha\Bigg)^2 + \Bigg(\sum_{\alpha=1}^{\nu} y'_\alpha\Bigg)\Bigg(\sum_{\beta=\nu+1}^{2\nu} x'_\beta\Bigg).$$

So, using

$$\int_{z\ge 0,\ \sum_{j=1}^m z_j\le a}\ \prod_{j=1}^m dz_j = \frac{a^m}{m!}, \qquad a\ge 0, \tag{3.30}$$

we have

$$\int_{D^*} \tilde{\Pi}(x,y)\,dx\,dy = \iiint_{0\le \xi_i<\infty} \exp\left[-\frac{1}{2}(\xi_1+\xi_2+\xi_3)^2+\xi_2\xi_3\right] \times \frac{\xi_1^{n-1}}{(n-1)!}\frac{\xi_2^{\nu-1}}{(\nu-1)!}\frac{\xi_3^{\nu-1}}{(\nu-1)!}d\xi_1\,d\xi_2\,d\xi_3, \tag{3.31}$$

where

$$\xi_1 = \sum_{\alpha=1}^{\nu} x_\alpha + \sum_{\alpha=2\nu+1}^{n} x_\alpha + \sum_{\alpha=\nu+1}^{2\nu} y_\alpha, \qquad \xi_2 = \sum_{\alpha=1}^{\nu} y'_\alpha, \qquad \xi_3 = \sum_{\alpha=\nu+1}^{2\nu} x'_\alpha.$$

Using $\exp(\xi_2\xi_3) = \sum_{k\ge 0}\xi_2^k\xi_3^k/k!$ and a formula

$$\overbrace{\int\cdots\int}^{m}_{0\le\xi_i<\infty} \exp\left[-\tfrac{1}{2}\left(\sum_{i=1}^{m}\xi_i\right)^2\right]\prod_{j=1}^{m}(\xi_j^{\alpha_j}\,d\xi_j) = \prod_{j=1}^{m}(\alpha_j!)\Bigg/\left(\sum_{i=1}^{m}(\alpha_i+1)-1\right)!! \tag{3.32}$$

[$\alpha_1,\dots,\alpha_m$ being integers such that $\sum_{i=1}^{m}(\alpha_i+1)$ is even], we finally get from (3.30):

$$\int_{D^*}\tilde{\Pi}(x,y)\,dx\,dy = \frac{1}{((\nu-1)!)^2}\sum_{k\ge 0}\frac{((\nu-1+k)!)^2}{k!(n+2\nu+2k-1)!!}. \tag{3.33}$$

[A simple derivation of (3.32) is based on (3.30) and $\int_0^\infty \exp(-s^2/2)s^l\,ds = (l-1)!!$, if l is odd.]

According to the last formula and (3.29), the value of $\int_{D_1}\tilde{\Pi}\,dx\,dy$ is bounded by a number which depends only on $\nu=\nu(G)$ and $\mu=\mu(G)$, the length of all circuits in G and the total number of circuits in G, respectively. We also notice that a given graph G [which is a collection of μ disjoint circuits of even lengths, and of total length 2ν, plus $(n-2\nu)/2$ isolated edges] gives rise to exactly 2^μ ordered pairs (M_1, M_2) such that $G(M_1,M_2)=G$, because for each circuit there are two ways to assign its edges alternately to M_1 and to M_2. Keeping these facts in mind, and using (3.28), (3.29) and (3.33), we obtain

$$\Sigma_n \le \exp(3n^{1/2}\log n)\sum_{\nu\ge\nu_1}\binom{n}{2\nu}(n-2\nu-1)!!\left(\sum_{\mu}2^{2\mu}f(2\nu,\mu)\right) \times\left(\sum_{k\ge 0}s(n,\nu,k)\right), \tag{3.34}$$

where

$$s(n,\nu,k) = \frac{((\nu-1+k)!)^2}{((\nu-1)!)^2 k!(n+2\nu+2k-1)!!} \tag{3.35}$$

and $f(2\nu, \mu)$ is the total number of ways to arrange a 2ν-element set in μ disjoint circuits, each of even length. [$\binom{n}{2\nu}(2n-2\nu-1)!!$ is the total number of ways to choose $n-2\nu$ members out of n, and arrange them in pairs.] We show in the Appendix that

$$\lim_{\nu\to\infty}((2\nu)!)^{-1}\sum_{\mu}2^{2\mu}f(2\nu,\mu)=e^{-1}, \tag{3.36}$$

so, in particular,

$$\sum_{\mu}2^{2\mu}f(2\nu,\mu)=O((2\nu)!). \tag{3.37}$$

Let us see what follows from the bound (3.34). First of all,

$$\frac{s(n,\nu,k+1)}{s(n,\nu,k)}=\frac{(\nu+k)^2}{(k+1)(n+2\nu+2k)}\le\frac{3}{4},\quad \text{if } k\ge 2\nu,$$

so that

$$\sum_{k\ge 0}s(n,\nu,k)=O\Big(\nu\max_{k\le 2\nu}s(n,\nu,k)\Big). \tag{3.38}$$

Applying Stirling's formula, we obtain: For $k\le 2\nu$,

$$s(n,\nu,k)=O(\exp(H(\nu,k))),$$

where

$$\begin{aligned}H(\nu,k)&=\frac{n}{2}+\nu-2\nu\log\nu+2(\nu+k)\log(\nu+k)-k\log k\\&\quad-\left(\frac{n}{2}+\nu+k\right)\log(n+2\nu+2k).\end{aligned} \tag{3.39}$$

As a function of a continuous argument k, $H(\nu,k)$ achieves its maximum at the root $\bar{k}$ of the equation

$$H_k'(\nu,k)=2\log(\nu+k)-\log k-\log(n+2\nu+2k)=0, \tag{3.40}$$

which is given by

$$\begin{aligned}\bar{k}=k(\nu)&=\nu^2\left[n/2+\sqrt{(n/2)^2+\nu^2}\right]^{-1}\\&\le\nu^2/n.\end{aligned} \tag{3.41}$$

If ν also varies continuously, then [by (3.39)–(3.41)]

$$\begin{aligned}\frac{d}{d\nu}H(\nu,k(\nu))&=\frac{\partial}{\partial\nu}H(\nu,k)\Big|_{k=\bar{k}}\\&=-2\log\nu+2\log(\nu+\bar{k})-\log(2n+2\nu+2\bar{k})\\&=\log\bar{k}-2\log\nu\\&=-\log n+\log\left(\frac{\bar{k}n}{\nu^2}\right)\le-\log n.\end{aligned} \tag{3.42}$$

Since

$$H(0,0) = \frac{n}{2}\log\frac{e}{n},$$

the inequality (3.42) and (3.38) imply that

$$\sum_{k\geq 0} s(n,\nu,k) = O\left(n\left(\frac{e}{n}\right)^{n/2} n^{-\nu}\right), \qquad \nu \geq 1. \tag{3.43}$$

Furthermore, according to (3.36),

$$\begin{aligned}
&\binom{n}{2\nu}(n-2\nu-1)!!\left(\sum_{\mu} 2^{2\mu} f(2\nu,\mu)\right)\\
&\quad = O\left((n-1)!!\prod_{j=0}^{\nu-1}(n-2j)\right) = O\left((n-1)!!n^{\nu}\exp\left(-\frac{\nu^2}{n}\right)\right)\\
&\quad = O\left(\left(\frac{n}{e}\right)^{n/2} n^{\nu}\log\left(\frac{-\nu^2}{n}\right)\right).
\end{aligned} \tag{3.44}$$

Combining (3.34), (3.43) and (3.44), we conclude that, since $\nu_1 = n^{3/4}\log n$,

$$\begin{aligned}
\Sigma_n &= O\left(n\exp(3n^{1/2}\log n)\sum_{\nu\geq\nu_1}\exp\left(-\frac{\nu^2}{n}\right)\right)\\
&= O\big(n^2\exp(3n^{1/2}\log n - n^{1/2}\log^2 n)\big)\\
&= O\big(\exp(-cn^{1/2}\log n)\big), \qquad \forall\, c > 0.
\end{aligned} \tag{3.45}$$

The relation (3.25) is established.

REMARK. The computation in (3.45) demonstrates that we had to choose ν_1 as large as we did in order to cancel this huge factor $\exp(3n^{1/2}\log n)$, which was obtained as an obvious bound for $\exp((3/2)\Sigma_{i=1}^n(x_i^2 + y_i^2))$. Now that we can concentrate on $\nu \leq \nu_1$, it becomes possible to discard all (x,y)'s for which this function exceeds $\exp(c\log^2 n)$ (Step 3) and then to reduce the range of dominant ν's to $\nu \leq 4n^{1/2}\log n$ (Step 4). Conceptually, this kind of reasoning resembles a bootstrapping technique which is used so effectively in asymptotic analysis; see de Bruijn [5], and Graham, Knuth and Patashnik [9], for instance.

STEP 3. Denoting $t_1 = \Sigma_{i=1}^n v_i^2$, $v = x/s$ and $t_{1*} = \Sigma_{i=1}^n v_{i*}^2$, $v_* = y/s_*$, consider

$$\Sigma_n' = \sum_{\nu\leq\nu_1}\int_{D_1\cap\{t_1\geq 3n^{-1}\}}\Pi(x,y)\,dx\,dy + \sum_{\nu\leq\nu_1}\int_{D_1\cap\{t_{1*}\geq 3n^{-1}\}}\Pi(x,y)\,dx\,dy.$$

According to (3.27), on D_1

$$\Pi(x,y) \leq \exp(3n^{1/2}\log n)\big(\exp(-s^2/2)\wedge\exp(-s_*^2/2)\big).$$

So (cf. Step 2),

$$\Sigma'_n \le 2\exp(3n^{1/2}\log n)$$
$$\times\left(\sum_{\nu\le\nu_1}\binom{n}{2\nu}(n-2\nu-1)!!\left(\sum_{\mu}2^{2\mu}f(2\nu,\mu)\right)\right)\int_{\{t_1\ge 3n^{-1}\}}\exp(-s^2/2)\,dx.$$

Here, according to (3.44), the sum is of order

$$(n-1)!!\sum_{\nu\le\nu_1}n^{\nu}=O\big((n-1)!!\exp(n^{3/4}\log^2 n)\big).$$

As for the integral, by Lemma 3 it is bounded by

$$\frac{(n-2)!!}{(n-1)!!}P\big(t_1(L^{(n)})\ge 3n^{-1}\big)=O\left(\frac{(n-2)!!}{(n-1)!!}e^{-cn}\right),\qquad c>0;$$

[the exponential bound for the probability is based on (3.4) and Lemma 4; compare with the bound (3.16)]. Thus,

$$\Sigma'_n=O\big(\exp(n^{3/4}\log^2 n-cn)\big),$$

which is exponentially small.

Consequently, (3.26) becomes

$$(3.46)\quad E(S_n(S_n-1))\sum_{\nu\le\nu_1}\int_{D_2}\Pi(x,y)\,dx\,dy+O\big(\exp(-cn^{1/2}\log n)\big),$$

where $D_2=D_2(M_1,M_2)\subset D_1$ is such that, in addition, $t_1,t_{1*}\le 3n^{-1}$.

STEP 4. For $(x,y)\in D_2$, the inequality (3.27) yields

$$(3.47)\qquad \Pi(x,y)\le\exp(9\log^2 n)\tilde{\Pi}(x,y).$$

Set $\nu_2=4n^{1/2}\log n$. Repeating verbatim the argument in Step 2,

$$\sum_{\nu\ge\nu_2}\int_{D_2}\Pi(x,y)\,dx\,dy=O\left(n\exp(9\log^2 n)\sum_{\nu\ge\nu_2}\exp(-\nu^2/n)\right)$$
$$=O\big(n^2\exp(-7\log^2 n)\big).$$

Thus [see (3.46)],

$$E(S_n(S_n-1))=\sum_{\nu\le\nu_2}\int_{D_2}\Pi(x,y)\,dx\,dy+O\big(\exp(-6\log^2 n)\big).$$

To repeat, D_2 is a subset of D such that

$$s,s_*\le n^{1/2}\log n,$$
$$t_1,t_{1*}\le 3n^{-1}$$

and

$$\nu_2=4n^{1/2}\log n.$$

STEP 5. As we shall see shortly, the dominant values of s, s_* and ν are all of exact order $n^{1/2}$, so that the bounds above are quite accurate. However, D_2 is still too "thick" for sharp asymptotic evaluations. A subset $D_3 = D_3(M_1, M_2) \subset D_2$, which will do the job, is defined by the following extra conditions:

$$\begin{aligned} &\sum_{i\in\mathscr{C}} v_i, \quad \sum_{i\in\mathscr{C}} v_{i*} \le n^{-1/2}\log^3 n,\\ &\sum_{i\in\mathscr{C}} v_i^2, \quad \sum_{i\in\mathscr{C}} v_{i*}^2 \le n^{-3/2}\log^3 n, \end{aligned} \tag{3.48}$$

and (cf. the proof of Theorem 1, Step 2),

$$\begin{aligned} &2n^{-1}(1-\varepsilon_n) \le \sum_{i=1}^n v_i^2, \quad \sum_{i=1}^n v_{i*}^2 \le 2n^{-1}(1+\varepsilon_n),\\ &2^{-1}n^{-1}(1-\varepsilon_n) \le \sum_{\{i,j\}\in M_1} v_i v_j, \quad \sum_{\{i,j\}\in M_2} v_{i*}v_{j*} \le 2^{-1}n^{-1}(1+\varepsilon_n),\\ &\sum_{i=1}^n v_i^4, \quad \sum_{i=1}^n v_{i*}^4 \le 25n^{-3}, \end{aligned} \tag{3.49}$$

where $\varepsilon_n = n^{-1/4}\log^2 n$.

Now, using (3.49) and $\tilde{\Pi}(x,y) \le \exp(-s^2/2)$, we estimate (via Lemma 3)

$$\begin{aligned} &\sum_{\nu\le\nu_2}\int_{D_2\cap\neg(3.49)} \Pi(x,y)\,dx\,dy\\ &\quad = O\left(\exp(9\log^2 n)\left(\sum_{\nu\le\nu_2} n^{\nu}(n-1)!!\right)\cdot\int_{\neg(3.49)}\exp(-s^2/2)\,dx\right)\\ &\quad = O\big(\exp(9\log^2 n + 4n^{1/2}\log^2 n)(Q_1+Q_2+Q_3)\big). \end{aligned}$$

Here

$$Q_1 = P\left(\left|\frac{t_1(L^{(n)})}{2n^{-1}} - 1\right| > \varepsilon_n\right),$$

$$Q_2 = P\left(\left|\frac{t_2(L^{(n)})}{2^{-1}n^{-1}} - 1\right| > \varepsilon_n\right)$$

and

$$Q_3 = P\left(\frac{t_3(L^{(n)})}{24n^{-3}} > \frac{25}{24}\right).$$

As we recall, Q_3 is exponentially small. Arguing as in the derivation of (3.16) [but with $n^{-1/4}\log n$ in (3.15) being replaced by $n^{-1/4}\log^{3/2} n$, which is still

$o(\varepsilon_n)$], we obtain

$$Q_1 = O\big(\exp\big(-c_1 n^{1/2}\log^3 n\big)\big), \qquad c_1 > 0.$$

Likewise,

$$Q_2 = O\big(\exp\big(-c_2 n^{1/2}\log^3 n\big)\big), \qquad c_2 > 0.$$

Therefore,

$$\sum_{\nu\le\nu_2}\int_{D_2\cap\neg(3.49)}\Pi(x,y)\,dx\,dy = O\big(\exp\big(-c_3 n^{1/2}\log^3 n\big)\big), \qquad c_3 > 0.$$

The same technique yields

$$\sum_{\nu\le\nu_2}\int_{D_2\cap\neg(3.48)}\Pi(x,y)\,dx\,dy = O\big(\exp\big(-c_4 n^{1/4}\log^3 n\big)\big), \qquad c_4 > 0.$$

Summarizing,

$$(3.50)\quad E(S_n(S_n-1)) = \sum_{\nu\le\nu_2}\int_{D_3}\Pi(x,y)\,dx\,dy + O\big(\exp\big(-6\log^2 n\big)\big).$$

REMARK. On the set D_3, the sums s, s_* and s_{**} are all close to each other, namely

$$(3.51)\qquad s/s_{**},\, s_*/s_{**} = 1 + O\big(n^{-1/2}\log^3 n\big).$$

STEP 6. Let us show that, uniformly over $\nu \le \nu_2$,

$$(3.52)\qquad \int_{D_3}\Pi(x,y)\,dx\,dy \le (1+o(1))2^{\mu}(\pi e/n)^{1/2}\frac{(n/e)^{n/2}}{n^{\nu}(n-1)!},$$

where $\mu = \mu(G)$ is the number of circuits of the graph $G = G(M_1, M_2)$.

To this end, we define $z_i = x_i \wedge y_i$, $1 \le i \le n$, and write

$$\Pi(x,y) = \prod_{\{i,j\}\in M_1^c\cap M_2^c}(1 - x_ix_j - y_iy_j + z_{ij})$$

$$(3.53)\qquad \le \exp\Bigg[-\sum_{\{i,j\}\in M_1^c\cap M_2^c}(x_ix_j + y_iy_j - z_iz_j) - \frac{1}{2}\sum_{\{i,j\}\in M_1^c\cap M_1^c}(x_ix_j + y_iy_j - z_iz_j)^2\Bigg] = \exp(-\Sigma_1 - \Sigma_2).$$

Now

$$\sum_{\{i,j\}\in M_1^c\cap M_2^c} a_ia_j = \frac{1}{2}\left(\sum_{i=1}^n a_i\right)^2 - \frac{1}{2}\left(\sum_{i=1}^n a_i^2\right) - \sum_{\{i,j\}\in M_1\cap M_2} a_ia_j - \left(\sum_{\{i,j\}\in M_1^c\cap M_2} a_ia_j + \sum_{\{i,j\}\in M_1\cap M_2^c} a_ia_j\right),$$

and the total value of the two last sums is at most $\sum_{i\in\mathscr{C}} a_i^2$, while

$$-\tfrac{1}{2}\sum_{i\in\mathscr{C}} a_i^2 \le \sum_{\{i,j\}\in M_1\cap M_2} a_i a_j - \sum_{\{i,j\}\in M_\alpha} a_i a_j \le \tfrac{1}{2}\sum_{i\in\mathscr{C}} a_i^2, \qquad \alpha = 1,2.$$

Consequently,

$$\Sigma_1 = \tfrac{1}{2}s^2(1 - t_1 - 2t_2) + \tfrac{1}{2}s_*^2(1 - t_{1*} - 2t_{2*})$$
$$- \tfrac{1}{2}s_{**}^2(1 - t_{1**} - 2t_{2**}) + O\Big(\sum_{i\in\mathscr{C}}(x_i^2 + y_i^2)\Big),$$

where, by definition, $t_{\alpha *} = t_\alpha(v_*)$, $t_{\alpha **} = t_\alpha(v_{**})$, $v_{**} = z/s_{**}$, $\alpha = 1,2$. Using (3.48), (3.49) and (3.51), we obtain from the last relation that on D_3,

$$\Sigma_1 = \frac{1}{2}s^2 + \frac{1}{2}s_*^2 - \frac{1}{2}s_{**}^2 - \frac{3}{2n}s_{**}^2 + O(n^{-1/4}\log^4 n). \tag{3.54}$$

Furthermore, Σ_2 is obviously close to $(1/2)\sum_{(i,j)} x_i^2 x_j^2 = (1/4)(\sum_i x_i^2)^2$. A careful check shows, via

$$(x_i x_j + y_i y_j - z_i z_j)^2 \le (x_i^2 + y_i^2)(x_j^2 + y_j^2),$$

that

$$\Sigma_2 = \frac{1}{4}\Big(\sum_{i=1}^n z_i^2\Big)^2 + O(n^{-1/2}\log^7 n)$$
$$= \frac{s_{**}^4}{n^2} + O(n^{-1/4}\log^6 n) + O(n^{-1/2}\log^7 n) \tag{3.55}$$
$$= \frac{s_{**}^4}{n^2} + O(n^{-1/4}\log^6 n).$$

[We could have written $(3/2n)s^2$, say, in (3.54) instead of $(3/2n)s_{**}^2$, and $-s^4/n^2$ instead of s_{**}^4/n^2 in (3.55). The benefits of our choice will become clear shortly.]

Thus, from (3.53) it follows that

$$\Pi(x,y) \le \exp\Big(-\frac{1}{2}s^2 - \frac{1}{2}s_*^2 + \frac{1}{2}s_{**}^2 + \frac{3}{2n}s_{**}^2 - \frac{1}{n^2}s_{**}^4$$
$$+ O(n^{-1/4}\log^6 n)\Big).$$

So, just like $\tilde{\Pi}(x,y)$ in Step 2, the bound above depends only on the sums s, s_* and s_{**}, if we neglect the remainder term. Furthermore, in the notation of

Step 2, (3.31), we have $s_{**} = \xi_1$, which is the promised benefit. Thus, similarly to (3.29) and (3.31), neglecting the factor $1 + O(n^{-1/4}\log^6 n)$,

$$\int_{D_3} \Pi(x,y)\,dx\,dy$$

$$\leq 2^{\mu(G)} \iiint_{\Delta} \exp\left[-\frac{1}{2}(\xi_1+\xi_2+\xi_3)^2 + \xi_2\xi_3 + \frac{3}{2n}\xi_1^2 - \frac{1}{n^2}\xi_1^4\right] \times \frac{\xi_1^{n-1}}{(n-1)!}\frac{\xi_2^{\nu-1}}{(\nu-1)!}\frac{\xi_3^{\nu-1}}{(\nu-1)!}\,d\xi_1\,d\xi_2,\xi_3. \tag{3.56}$$

The set Δ, according to the definition of D_3 and the genesis of the variables ξ_1,ξ_2,ξ_3, is given by

$$\begin{aligned} &0 \leq \xi_1 \leq n^{1/2}\log n,\\ &0 \leq \xi_2, \xi_3 \leq \log^4 n. \end{aligned} \tag{3.57}$$

Denote the corresponding integral by I_ν.

Given ξ_2, ξ_3, we estimate first

$$\begin{aligned} &\int_0^\infty \exp\left[-\frac{1}{2}(\xi_1+\xi_2+\xi_3)^2 + \frac{3}{2n}\xi_1^2 - \frac{1}{n^2}\xi_1^4\right]\xi_1^{n-1}\,d\xi_1\\ &\quad= \int_0^\infty \xi_1^{-1}\exp\left(\frac{3}{2n}\xi_1^2 - \frac{1}{n^2}\xi_1^4\right)\exp(H(\xi_1,\xi_2,\xi_3))\,d\xi_1, \end{aligned} \tag{3.58}$$

where

$$H(\xi_1,\xi_2,\xi_3) = -\tfrac{1}{2}(\xi_1+\xi_2+\xi_3)^2 + n\log\xi_1.$$

As a function of ξ_1, H is concave down, and it achieves its maximum at

$$\begin{aligned} \bar{\xi}_1 &= \frac{1}{2}\left(-\eta + (\eta^2+4n)^{1/2}\right) \qquad (\eta =_{\mathrm{def}} \xi_2+\xi_3)\\ &= \sqrt{n} - \frac{1}{2}\eta + \frac{n^2}{8n^{1/2}} + O(n^{-3/2}\log^{16} n). \end{aligned} \tag{3.59}$$

A direct computation shows that

$$\begin{aligned} H(\bar{\xi}_1,\xi_2,\xi_3) &= \frac{n}{2}\log\left(\frac{n}{e}\right) - \sqrt{n}(\xi_2+\xi_3) - \frac{1}{4}(\xi_2+\xi_3)^2\\ &\quad + O(n^{-1}\log^{16} n). \end{aligned}$$

Also,

$$H_{\xi_1}^{(2)}(\bar{\xi}_1,\xi_2,\xi_3) = -1 - \frac{n}{\bar{\xi}_1^2} = -2 + O(n^{-1/2}\log^4 n),$$

so (Laplace's method) the integral in (3.58) is

$$
\begin{aligned}
&(1+o(1))n^{-1/2}\exp\left(\frac{1}{2}+H(\bar{\xi}_1,\xi_2,\xi_3)\right)\left(\frac{2\pi}{2}\right)^{1/2}\\
(3.60)\qquad &=(1+o(1))\left(\frac{\pi e}{n}\right)^{1/2}\left(\frac{n}{e}\right)^{n/2}\exp\left[-\sqrt{n}\,(\xi_2+\xi_3)-\frac{1}{4}(\xi_2+\xi_3)^2\right].
\end{aligned}
$$

Next,

$$
\begin{aligned}
&\int_0^\infty\int_0^\infty \exp\left[-\sqrt{n}\,(\xi_2+\xi_3)-\tfrac{1}{4}(\xi_2+\xi_3)^2+\xi_2\xi_3\right]\xi_2^{\nu-1}\xi_3^{\nu-1}\,d\xi_2\,d\xi_3\\
(3.61)\qquad &=\int_0^\infty\int_0^\infty \exp\left[-\sqrt{n}\,(\xi_2+\xi_3)-\tfrac{1}{4}(\xi_2-\xi_3)^2\right]\xi_2^{\nu-1}\xi_3^{\nu-1}\,d\xi_2\,d\xi_3\\
&\le\left(\int_0^\infty \exp(-\sqrt{n}\,\xi)\xi^{\nu-1}\,d\xi\right)^2=n^{-\nu}((\nu-1)!)^2.
\end{aligned}
$$

Thus [see (3.60)],

$$
(3.62)\qquad I_\nu\le(1+o(1))\left(\frac{\pi e}{n}\right)^{1/2}\frac{(n/e)^{n/2}}{n^\nu(n-1)!},
$$

and, by (3.56), the bound (3.52) follows.

This estimate and the formula (3.50) show that

$$
\begin{aligned}
&E(S_n(S_n-1))\\
(3.63)\qquad &\le(1+o(1))\left(\frac{\pi e}{n}\right)^{1/2}\frac{(n/e)^{n/2}}{(n-1)!}\sum_{\nu\le\nu_2}n^{-\nu}\binom{n}{2\nu}(n-2\nu-1)!!\\
&\cdot\left(\sum_\mu 2^{2\mu}f(2\nu,\mu)\right)+O\big(\exp(-6\log^2 n)\big).
\end{aligned}
$$

In view of (3.36), the last sum is asymptotic to

$$
\begin{aligned}
&e^{-1}(n-1)!!\sum_{\nu\le\nu_2}n^{-\nu}\prod_{j=0}^{\nu-1}(n-2j)\\
(3.64)\qquad &=e^{-1}(n-1)!!\sum_{\nu\le\nu_2}\exp\big(-\nu^2/n+O(\nu_2^3/n^2)\big)\\
&=(1+o(1))e^{-1}(n-1)!!(\pi n)^{1/2}/2,
\end{aligned}
$$

($\int_0^\infty e^{-x^2}\,dx=\sqrt{\pi}/2$). Plugging (3.64) into (3.63), and using

$$
\begin{aligned}
\frac{(n-1)!!}{(n-1)!}&=((n-2)!!)^{-1}=\left(2^{(n-2)/2}\left(\frac{n-2}{2}\right)!\right)^{-1}\\
&=\left(1+O\left(\frac{1}{n}\right)\right)e(\pi n)^{-1/2}\left(\frac{e}{n}\right)^{(n/2)-1},
\end{aligned}
$$

we conclude that

$$E(S_n(S_n-1)) \le (1+o(1))\left(\frac{\pi n}{4e}\right)^{1/2}. \tag{3.65}$$

REMARK. We should notice that the double integral (3.61) is, in fact, asymptotic to $n^{-\nu}((\nu-1)!)^2$, since the integrand achieves its sharply pronounced maximum at a point $(\bar{\xi}_2, \bar{\xi}_3) = ((\nu-1)/\sqrt{n}, (\nu-1)/\sqrt{n})$, for which $(\xi_2 - \xi_3)^2 = 0$. Also, since $\nu \le 4n^{1/2}\log n$,

$$\bar{\xi}_2, \bar{\xi}_3 \le 4\log n,$$

and [see (3.59)],

$$\bar{\xi}_1 = \bar{\xi}_1(\xi_2, \xi_3) = \sqrt{n} + O(\log^4 n).$$

In view of the conditions (3.57), we can see then that I_ν is asymptotic to its upper bound given in (3.62).

STEP 7. The insight we have gained makes it relatively easy to demonstrate that

$$E(S_n(S_n-1)) \ge (1+o(1))\left(\frac{\pi n}{4e}\right)^{1/2},$$

as well. All we have to do is to confine ourselves to the summands, and the integration domains, which we expect to contribute most to the value of $E(S_n(S_n-1))$, and then to estimate sharply their overall contribution.

Define $\nu_{\pm} = n^{1/2}\log^{\pm 1} n$. For $\nu_- \le \nu \le \nu_+$, introduce $D_4 = D_4(M_1, M_2)$, a subset of $D = D(M_1, M_2)$, such that, in addition,

$$x_i, y_i \le \delta_n =_{\text{def}} \frac{4\log^3 n}{n^{1/2}}, \qquad 1 \le i \le n, \tag{3.66}$$

$$n^{1/2}(1-\varepsilon_n) \le s_0 =_{\text{def}} \sum_{i\in\mathscr{C}^c} x_i \le n^{1/2}(1+\varepsilon_n), \qquad (\mathscr{C} = \mathscr{C}(M_1, M_2)), \tag{3.67}$$

and, denoting $u_i = x_i/s_0$, $i \in \mathscr{C}^c$,

$$2n^{-1}(1-\varepsilon_n) \le \sum_{i\in\mathscr{C}^c} u_i^2 \le 2n^{-1}(1+\varepsilon_n),$$

$$2^{-1}n^{-1}(1-\varepsilon_n) \le \sum_{\{i,j\}\in M_1\cap M_2} u_i u_j \le 2^{-1}n^{-1}(1+\varepsilon_n).$$

Here $\varepsilon_n = n^{-1/4}\log^2 n$.

Using $1-\gamma = \exp(-\gamma - \gamma^2/2 + O(\gamma^3))$, $\gamma \to 0$, and the condition (3.66), we get

$$\begin{aligned}\Pi(x,y) &= \exp\left[-\Sigma_1 - \Sigma_2 + O\left(\left(\sum_{i=1}^{n}(x_i \vee y_i)^3\right)^2\right)\right]\\ &= \exp\left[-\Sigma_1 - \Sigma_2 + O(\log^{18} n/n)\right];\end{aligned}$$

see (3.53) for the definition of the sums Σ_1, Σ_2. Strengthening [via (3.66) and (3.67)] the argument in Step 6, we obtain further that

$$\Pi(x,y) = \exp\left(\tfrac{1}{2} + O(\varepsilon_n)\right)\tilde{\Pi}(x,y),$$
$$\tilde{\Pi}(x,y) = \exp\left(-\tfrac{1}{2}s^2 - \tfrac{1}{2}s_*^2 + \tfrac{1}{2}s_{**}^2\right)$$

[$e^{1/2}$ is the asymptotic value of $\exp((3/2n)s_0^2 - (1/n^2)s_0^4)$]. So, within the factor $1 + O(\varepsilon_n)$,

$$\int_{D_4} \Pi(x,y)\,dx\,dy = 2^{\mu(G)} e^{1/2} \int_{D^\wedge} \tilde{\Pi}(x,y)\left(\prod_{\alpha=1}^{n} dx_\alpha \prod_{\beta=1}^{2\nu} dy_\beta\right), \tag{3.68}$$

when $D^\wedge$ is such that

$$\delta_n \ge y_\alpha \ge x_\alpha \ge 0, \quad 1 \le \alpha \le \nu, \tag{3.69}$$
$$\delta_n \ge x_\alpha \ge y_\alpha \ge 0, \quad \nu + 1 \le \alpha \le 2\nu,$$

$$\delta_n \ge x_\alpha \ge 0, \quad 2\nu + 1 \le \alpha \le n, \tag{3.70}$$

$$n^{1/2}(1 - \varepsilon_n) \le s_0 = \sum_{i=2\nu+1}^{n} x_i \le n^{1/2}(1 + \varepsilon_n), \tag{3.71}$$

$$2n^{-1}(1 - \varepsilon_n) \le \sum_{i=2\nu+1}^{n} u_i^2 \le 2n^{-1}(1 + \varepsilon_1), \tag{3.72}$$

$$2^{-1}n^{-1}(1 - \varepsilon_n) \le \sum_{i=2\nu+1}^{(n+2\nu)/2} u_i u_{i+(n-2\nu)/2} \le 2^{-1}n^{-1}(1 + \varepsilon_n), \tag{3.73}$$

$u_i = x_i/s_0$, $2\nu + 1 \le i \le n$. In the variables (x', y') (see Step 2), the conditions (3.70) remain unchanged, while (3.69) becomes

$$\delta_n \ge x_\alpha \ge 0, \quad y'_\alpha \ge 0, \quad x_\alpha + y'_\alpha \le \delta_n, \quad 1 \le \alpha \le \nu,$$
$$\delta_n \ge y_\alpha \ge 0, \quad x'_\alpha \ge 0, \quad x'_\alpha + y_\alpha \le \delta_n, \quad \nu + 1 \le \alpha \le 2\nu.$$

The latter are certainly satisfied if

$$\begin{aligned} &0 \le x_\alpha, \quad y'_\alpha \le \delta_n/2, \quad 1 \le \alpha \le \nu, \\ &0 \le x'_\alpha, \quad y_\alpha \le \delta_n/2, \quad \nu + 1 \le \alpha \le 2\nu. \end{aligned} \tag{3.74}$$

Let us further reduce the domain of integration, by requiring that

$$\begin{aligned} &s_1 =_{\text{def}} \sum_{i=1}^{\nu} x_i, \qquad s_2 =_{\text{def}} \sum_{i=1}^{\nu} y'_i \le 2\log n, \\ &s_3 =_{\text{def}} \sum_{i=\nu+1}^{2\nu} x'_i, \qquad s_4 =_{\text{def}} \sum_{i=\nu+1}^{2\nu} y_i \le 2\log n. \end{aligned} \tag{3.75}$$

[In the notation of (3.31), $s_2 = \xi_2$, $s_3 = \xi_3$.] Two important properties of $\tilde{\Pi}(x,y)$ and the conditions (3.71)–(3.75) are straightforward.

(a) $\tilde{\Pi}(x,y)$ depends on (x,y) only through the sums s_i, $0 \le i \le 4$, namely [see the integrand in (3.31)]:

$$\tilde{\Pi}(x,y) = \exp\left[-\tfrac{1}{2}\left(\sum_{i=0}^{4} s_i\right)^2 + s_2 s_3\right].$$

(b) Each of the five groups of variables, $\{x_i;\ 2\nu + 1 \le i \le n\}$, $\{x_i;\ 1 \le i \le \nu\}$, $\{y_i';\ 1 \le i \le \nu\}$, $\{x_i';\ \nu + 1 \le i \le 2\nu\}$ and $\{y_i';\ \nu + 1 \le i \le 2\nu\}$, enters its own set of the conditions.

So, using Lemma 3, we can assert that

$$\int_{D^\wedge} \tilde{\Pi}\, dx\, dy \ge \overbrace{\int \cdots \int}^{5}_{s_0, \dots, s_4} \exp\left[-\frac{1}{2}\left(\sum_{i=0}^{4} s_i\right)^2 + s_2 s_3\right]$$
$$\times \frac{P_0(s_0) s_0^{n-2\nu-1}}{(n-2\nu-1)!} \prod_{i=1}^{4} \frac{P(s_i) s_i^{\nu-1}}{(\nu-1)!} (ds_0 \cdots ds_4).$$

Here s_i, $0 \le i \le 4$, satisfy (3.71) and (3.75); furthermore,

$$P_0(s_0) = P\Bigg(\max_{1 \le j \le n-2\nu} L_j^{(n-2\nu)} \le \delta_n / s_0;\ 2n^{-1}(1 - \varepsilon_n)$$
$$\le \sum_{j=1}^{n-2\nu} \left(L_j^{(n-2\nu)}\right)^2 \le 2n^{-1}(1 + \varepsilon_n);$$
$$2^{-1}n^{-1}(1 - \varepsilon_n) \le \sum_{j=1}^{(n-2\nu)/2} L_j^{(n-2\nu)} L_{j+(n-2\nu)/2}^{(n-2\nu)} \le 2^{-1}n^{-1}(1 + \varepsilon_n)\Bigg),$$

and

$$P(\theta) = P\left(\max_{1 \le j \le \nu} L_j^{(\nu)} \le \delta_n / 2\theta\right).$$

[$L_j^{(m)}(1 \le j \le m)$ are the lengths of m subintervals of [0, 1] induced by $(m - 1)$ random points.] Now, since $n^{1/2} \log^{-1} n \le \nu \le 4n^{1/2} \log n$, we have

$$\frac{\delta_n}{2s_i} \ge \frac{\delta_n}{4 \log n} = \frac{\log^2 n}{n^{1/2}}$$
$$= \frac{\log n}{\nu} = \left(2 + O\left(\frac{\log\log n}{\log n}\right)\right) \frac{\log \nu}{\nu}.$$

Thus, by (3.19),

$$\prod_{i=1}^{4} P(s_i) = 1 - O(\nu^{-1}).$$

Analogously (see the proof of Theorem 1), $P_0(s_0) \to 1$ uniformly over the integration domain, since

$$\frac{\delta_n}{s_0} \geq \frac{2\log^3 n}{n}(1+\varepsilon_n)^{-1} \gg \frac{\log n}{n},$$

and

$$n - 2\nu = n\big(1 + O(n^{-1/2}\log n)\big) = n\big(1 + o(\varepsilon_n)\big).$$

Consequently, within a factor approaching 1,

$$\int_{D^\wedge} \tilde{\Pi}\, dx\, dy \geq \frac{1}{(n-2\nu-1)!((\nu-1)!)^4} \overbrace{\int \cdots \int}^{5}_{s_i} \exp\left[-\frac{1}{2}\left(\sum_{i=0}^{4} s_i\right)^2 + s_2 s_3\right]$$

$$\times s_0^{n-2\nu-1}\left(\prod_{i=1}^{4} s_i^{\nu-1}\right) ds_0\, ds_1\, ds_2\, ds_3\, ds_4,$$

where

$$n^{1/2}(1-\varepsilon_n) \leq s_0 \leq n^{1/2}(1+\varepsilon_n), \tag{3.76}$$

$$0 \leq s_i \leq 2\log n, \qquad i = 1,2,3,4. \tag{3.77}$$

The rest resembles the computations in Step 6. We write the integrand as

$$\left(s_0^{-1}\prod_{i=1}^{4} s_i^{-1}\right)\exp[H(s)],$$

where

$$H(s) =_{\text{def}} -\tfrac{1}{2}\left(\sum_{i=0}^{4} s_i\right)^2 + s_2 s_3 + n'\log s_0 + \nu\sum_{i=1}^{4}\log s_i \qquad (n' = n - 2\nu)$$

is strictly concave down. Given that s_i, $i = 1,2,3,4$, satisfy (3.77), as a function of s_0, H achieves its maximum at

$$\bar{s}_0 = \sqrt{n'} - \frac{1}{2}\theta + \frac{\theta^2}{8\sqrt{n'}} + O(n^{-3/2}\log^4 n), \qquad \theta = \sum_{i=1}^{4} s_i.$$

So,

$$H(\bar{s}_0, s_1, s_2, s_3, s_4) = \frac{n'}{2}\log\left(\frac{n'}{e}\right) - \sqrt{n'}\,\theta - \frac{1}{4}\theta^2 + s_2 s_3 + O(n^{-1}\log^4 n)$$

and

$$H_{s_0}^{(2)}(\bar{s}_0, s_1, s_2, s_3, s_4) = -2 + O(n^{-1/2}\log n).$$

Besides, $|n^{1/2}(1 \pm \varepsilon_n) - \bar{s}_0|$ is of order at least $n^{1/2}\varepsilon_n = n^{1/4}\log^4 n$, since $n' = n(1 + O(\nu/n)) = n(1 + O(n^{-1/2}\log n))$. In other words, $\bar{s}_0$ is well within the boundaries (3.76). With the help of concavity of H, we obtain then that the integral with respect to s_0 is asymptotic to

$$
\begin{aligned}
&\left(\frac{n'}{e}\right)^{n'/2}\left(\prod_{i=1}^{4} s_i^{-1}\right)\bar{s}_0^{-1}\exp[\mathscr{H}(s)]\left(\frac{2\pi}{\left(-H_{s_0}^{(2)}(\bar{s}_0, s_1, s_2, s_3, s_4)\right)}\right)^{1/2} \\
(3.78)\qquad &= \left(\frac{\pi}{n}\right)^{1/2}\left(\frac{n'}{e}\right)^{n'/2}\left(\prod_{i=1}^{4} s_i^{-1}\right)\exp[\mathscr{H}(s)],
\end{aligned}
$$

where

$$\mathscr{H}(s) =_{\text{def}} -\sqrt{n'}\left(\sum_{i=1}^{4} s_i\right) - \tfrac{1}{4}\left(\sum_{i=1}^{4} s_i\right)^2 + s_2 s_3 + \nu \sum_{i=1}^{4} \log s_i.$$

In its turn, $\mathscr{H}$ is also strictly concave down, and its maximum is attained at $\bar{s} = (\bar{s}_1, \bar{s}_2, \bar{s}_3, \bar{s}_4)$ such that

$$\bar{s}_1 = \bar{s}_4 = \frac{\nu}{\sqrt{n'}}\left(1 - \frac{2\nu}{n'} + O\left(\frac{\nu}{n'}\right)^2\right),$$

$$\bar{s}_2 = \bar{s}_3 = \frac{\nu}{\sqrt{n'}}\left(1 - \frac{\nu}{n'} + O\left(\left(\frac{\nu}{n'}\right)^2\right)\right).$$

Since $\nu \le \nu_+ = n^{1/2}\log n$, $n' = n - 2\nu$, we see that $\bar{s}_i \le (1 + o(1))\log n$, $1 \le i \le 4$, that is $\bar{s}$ satisfies the restriction (3.77), and the distance from $\bar{s}$ to the boundary is of an exact order $\log n$. After some work, we obtain that

$$\mathscr{H}(\bar{s}) = 4\nu \log\left(\frac{\nu}{e}\right) - 2\nu \log n + \frac{\nu^2}{n} + O\left(\frac{\nu^3}{n^2}\right),$$

that is, the remainder term is $O(n^{-1/2}\log^3 n)$. Besides,

$$\left[\mathscr{H}''_{s_i s_j}(\bar{s})\right]_{i,j=1}^{4} = -\frac{n}{\nu} I_4 + O(1),$$

I_4 being the 4×4 identity matrix. We notice also that for every s on the boundary of the 4-cube (3.77), $\|s - \bar{s}\|$ is of order $\log^{-1} n$, at least, since $\nu \ge \nu_- = n^{1/2}\log^{-1} n$; so, $(s - \bar{s})^T\mathscr{H}''(\bar{s})(s - \bar{s})$ is of order $n^{1/2}\log^{-3} n$, at least. Then, integrating (3.78) over s_1, s_2, s_3, s_4, we can claim that within a factor $1 + o(1)$,

$$\int_{D^\wedge} \tilde{\Pi}\, dx\, dy \ge \frac{(\pi/n)^{1/2}(n'/e)^{n'/2}}{(n'-1)!((\nu-1)!)^4}\frac{n^2}{\nu^4}\left(\frac{\nu}{e}\right)^{4\nu}\frac{n^{-2\nu}e^{\nu^2/n}(2\pi)^2}{(n/\nu)^2},$$

$[\det[\mathscr{H}''(\bar{s})] \sim (n/\nu)^4]$. Now, $\nu! \sim (2\pi\nu)^{1/2}(\nu/e)^\nu$ and

$$\frac{(n'/e)^{n'/2}}{(n'-1)!} \sim \left(\frac{n}{2\pi}\right)^{1/2}\left(\frac{n}{e}\right)^{-n/2} n^\nu e^{-\nu^2/n}.$$

Hence, after massive cancellations,

$$\int_{D^{\wedge}} \tilde{\Pi}\, dx\, dy \geq (1 + o(1)) \frac{1}{\sqrt{2}} \left(\frac{n}{e} \right)^{-n/2} n^{-\nu}.$$

Then [see (3.68)],

$$\int_{D_4} \Pi(x, y)\, dx\, dy \geq (1 + o(1)) 2^{\mu(G)} \left(\frac{e}{2} \right)^{1/2} \left(\frac{n}{e} \right)^{-n/2} n^{-\nu},$$

and we observe that this lower bound is asymptotic to the upper bound in (3.52).

Thus, since $\nu_- / \sqrt{n} \to 0$, $\nu_+ / \sqrt{n} \to \infty$, using (3.36) we come to

$$\begin{aligned} E(S_n(S_n - 1)) &\geq (1 + o(1)) \left(\frac{e}{2} \right)^{1/2} \left(\frac{n}{e} \right)^{-n/2} \\ &\quad \times \sum_{\nu = \nu_-}^{\nu_+} n^{-\nu} \binom{n}{2\nu} (n - 2\nu - 1)!! \left(\sum_{\mu} 2^{2\mu} f(2\nu, \mu) \right) \\ &= (1 + o(1)) \left(\frac{n}{e} \right)^{1/2} \int_{\nu_-/\sqrt{n}}^{\nu_+/\sqrt{n}} e^{-x^2}\, dx \\ &= (1 + o(1)) \frac{(\pi n / e)^{1/2}}{2}. \end{aligned}$$

Taken together with (3.65), this finishes the proof of Theorem 2, since $E(S_n) = O(1)$. □

COROLLARY.

$$P(S_n \geq 1) \geq (1 + o(1)) \left(\frac{4e^3}{\pi n} \right)^{1/2}.$$

The proof is immediate, based on Theorems 1 and 2, and the inequality

$$P(|X| > 0) \geq E^2(|X|) / E(X^2).$$

(The last relation is a direct consequence of Cauchy's inequality.)

Consequently, if $P(S_n \geq 1)$ goes to 0, it does so quite reluctantly, not faster than $n^{-1/2}$. Considering how simple-minded our last argument is, we would venture to guess that the true rate of convergence is slower, something of the order $n^{-\sigma}$, $\sigma < 1/2$, or maybe even $(\log n)^{-\rho}$, $\rho > 0$. [Of course, a logical possibility remains that $P(S_n \geq 1)$ is bounded away from 0.]

[A reviewer's comment: "For a critical finite variance simple branching process $\{Z_n =$ the size of the nth generation$\}$, $E(Z_n) \equiv 1$, $E(Z_n^2) \sim Cn$ and $P(Z_n \geq 1) \sim C_1/n$, so that the corollary could well be sharp."]

4. Stable ranks. In this last section, we prove that the stable matchings, when they exist, are likely to be well balanced, in a sense that in every stable matching the partners are likely to be close to the top of each other's preference lists.

Here are the precise statements.

THEOREM 3. *For a given stable matching $\mathscr{M}$ let $\mathscr{L}(\mathscr{M})$ denote the largest rank of a partner. Then, for every $\varepsilon > 0$,*

$$P(\nexists\, \mathscr{M} \text{ s.t. } \mathscr{L}(\mathscr{M}) \le (1-\varepsilon)n^{1/2}\log n) = 1 - O(e^{-n^c}), \qquad \forall\, c < \varepsilon \wedge 1/2, \tag{4.1}$$

$$P(\nexists\, \mathscr{M} \text{ s.t. } \mathscr{L}(\mathscr{M}) \ge (1+\varepsilon)n^{1/2}\log n) = 1 - O(n^{-c}), \qquad \forall\, c < \varepsilon. \tag{4.2}$$

[In (4.1), $\varepsilon < 1$.] Consequently, $\mathscr{L}(\mathscr{M})$ is asymptotic, in probability, to $n^{1/2}\log n$, uniformly over all stable matchings $\mathscr{M}$.

THEOREM 4. *Let $R(\mathscr{M})$ denote the sum of the ranks of all partners in a stable matching $\mathscr{M}$. Then*

$$P(\nexists\, \mathscr{M} \text{ s.t. } R(\mathscr{M}) \le (1-\varepsilon)n^{3/2}) = 1 - O(e^{-cn}), \qquad \forall\, \varepsilon \in (0,1) \tag{4.3}$$

for every

$$c < -\log(1-\varepsilon) - \varepsilon + \varepsilon^2/2,$$

and

$$P(\nexists\, \mathscr{M} \text{ s.t. } R(\mathscr{M}) \ge (1+\varepsilon)n^{3/2}) = 1 - O(e^{-cn}), \qquad \forall\, \varepsilon > 0 \tag{4.4}$$

for every

$$c < -\log(1+\varepsilon) + \varepsilon + \varepsilon^2/2.$$

Consequently [the function $f(z) =_{\text{def}} -\log(1+z) + z + z^2/2 > 0$ for $z > -1$], $R(\mathscr{M})$ is asymptotic to $n^{3/2}$, uniformly over all stable matchings $\mathscr{M}$, with exponentially high probability.

REMARK. So (see the Corollary), given that at least one stable matching exists,

$$(1-\varepsilon)n^{1/2}\log n \le \mathscr{L}(\mathscr{M}),$$
$$(1-\varepsilon)n^{3/2} \le R(\mathscr{M}) \le (1+\varepsilon)n^{3/2}$$

with superpolynomially high probability, and

$$\mathscr{L}(\mathscr{M}) \le (1.5+\varepsilon)n^{1/2}\log n$$

with probability $\ge 1 - O(n^{-c})$, $\forall\, c < \varepsilon$, uniformly over all stable matchings.

PROOF OF THEOREM 3. The argument is based on the bounds

$$P\left(\max_j L_j^{(n)} \le (1-\varepsilon)\frac{\log n}{n}\right) = O(\exp(-n^c)), \qquad \forall\, c < \varepsilon,$$

([22], Appendix) and

$$(4.5)\qquad P\left(\max_j L_j^{(n)} \ge (1+\varepsilon)\frac{\log n}{n}\right) = O(n^{-c}), \qquad \forall\, c < \varepsilon.$$

[cf. (3.19)].

Consider, for instance, (4.2). For a stable matching $\mathscr{M}$, introduce $X(\mathscr{M}) =_{\text{def}} \max_i X_{i,m(i)}$, where $m(i)$ is the corresponding matching function.

LEMMA 5. *For every $\delta > 0$,*

$$P_n(\delta) =_{\text{def}} P\left(\nexists\, \mathscr{M}\ s.t.\ X(\mathscr{M}) \ge (1+\delta)\frac{\log n}{n^{1/2}}\right) = 1 - O(n^{-c}), \qquad \forall\, c < \delta.$$

PROOF. First of all,

$$1 - P_n(\delta) \le (n-1)!!\int_{\max_i x_i \ge (1+\delta)\log n/n^{1/2}} \Pi(x)\,dx,$$

where [see (3.5)]

$$\Pi(x) \le \exp\left(-\tfrac{1}{2}s^2 + \sum_{i=1}^n x_i^2\right).$$

Second, the quantity

$$(n-1)!!\int_{t_1 \ge 3/n} \Pi(x)\,dx, \quad \text{where} \quad t_1 = \sum_{i=1}^n (x_i/s)^2,$$

is $O(e^{-\alpha n})$, $\alpha > 0$ (see Steps 1 and 2 in the proof of Theorem 1). So, neglecting an exponentially small term,

$$1 - P_n(\delta) \le (n-1)!!\int_{\max x_i \ge (1+\delta)\log n/n^{1/2}} \exp\left(-\frac{1}{2}s^2\left(1 - \frac{6}{n}\right)\right)dx.$$

Fix $\delta_1 \in (0, \delta)$, and break the integration domain into two parts, $s \ge (1+\delta_1)n^{1/2}$ and $s < (1+\delta_1)n^{1/2}$. The contribution of the first domain is at most

$$\begin{aligned}
&\frac{(n-1)!!}{(n-1)!}\int_{(1+\delta_1)n^{1/2}}^{\infty} \exp\left(-\frac{1}{2}s^2\left(1-\frac{6}{n}\right)\right)s^{n-1}\,ds \\
&\quad = O\left(\frac{(n-1)!!}{(n-1)!}\exp\left(-\frac{1}{2}n(1+\delta_1)^2\right)\left((1+\delta_1)n^{1/2}\right)^n\right) \\
&\quad = O\left(\frac{(n-1)!!\,(n/e)^{n/2}}{(n-1)!}\exp(-nf(\delta_1))\right) = O(\exp(-nf(\delta_1))),
\end{aligned}$$

which is exponentially small. The second domain contributes at most (see

Lemma 3)

$$\frac{(n-1)!!}{(n-1)!}\left(\int_0^\infty \exp\left(-\frac{1}{2}s^2\left(1-\frac{6}{n}\right)\right)s^{n-1}\,ds\right)P\left(\max_i L_i^{(n)} \geq \frac{1+\delta}{1+\delta_1}\frac{\log n}{n}\right)$$

$$= \left(1-\frac{6}{n}\right)^{n/2} P\left(\max_i L_i^{(n)} \geq \frac{1+\delta}{1+\delta_1}\frac{\log n}{n}\right).$$

It remains to notice that, by (4.5), the probability factor is $O(n^{-c})$, if

$$c < \frac{\delta - \delta_1}{1+\delta_1}.$$

Clearly, c can be made arbitrarily close to δ if we choose $\delta_1 > 0$ sufficiently small. □

To continue, let us choose $\delta \in (0, \varepsilon)$. Recall that the variables X_{ij} are independent and uniform [0, 1]. Using Chernoff's bound for tails of the binomial distribution [3], one can easily show that a fixed row of $[X_{ij}]$ contains at most $(1+\varepsilon)n^{1/2}\log n$ entries not exceeding $(1+\delta)\log n/n^{1/2}$, with probability $\geq 1 - \exp(-c_1 n^{1/2}\log n)$, for every c_1 such that

$$c_1 < (1+\varepsilon)\log\frac{1+\varepsilon}{1+\delta} + \delta - \varepsilon.$$

Invoking Lemma 5, we obtain then

$$\begin{aligned} &P(\exists\, \mathscr{M} \text{ s.t. } \mathscr{L}(\mathscr{M}) \geq (1+\varepsilon)n^{1/2}\log n) \\ &\quad = O(n^{-c}) + P(\exists\, \mathscr{M} \text{ s.t. } \mathscr{L}(\mathscr{M}) \\ &\qquad\qquad \geq (1+\varepsilon)n^{1/2}\log n,\ X(\mathscr{M}) < (1+\delta)\log n/n^{1/2}) \\ &\quad = O(n^{-c}) + O\big(n\exp(-c_1 n^{1/2}\log n)\big) \\ &\quad = O(n^{-c}). \end{aligned}$$

(Recall that, by definition, $X(\mathscr{M}) = \max_i X_{i,m(i)}$, and

$$\mathscr{L}(\mathscr{M}) = \max_i |\{j \neq i \colon X_{ij} < X_{i,m(i)}\}| + 1.\Big)$$

The relation (4.2) follows, since we can select δ arbitrarily close to ε, and c arbitrarily close to δ. □

Finally, we have the following:

PROOF OF THEOREM 4. We prove only (4.4), since the argument for (4.3) is similar. Denote $\bar{k} = [(1+\varepsilon)n^{3/2}]$. We have obviously

$$P\big(\exists\, \mathscr{M} \text{ s.t. } R(\mathscr{M}) \geq \bar{k}\big) \leq (n-1)!! \sum_{k \geq \bar{k}} P_{nk},$$

where P_{nk} is the probability that the standard matching M_0 is stable and its total partners' rank equals k. By Lemma 1 (2.2),

$$\sum_{k\geq\bar{k}} P_{nk} = \int_C \sum_{k>\bar{k}} [z^{k-n}] \prod_{\{i,j\}\in M_0^c} \left(\bar{x}_i\bar{x}_j + zx_i\bar{x}_j + z\bar{x}_ix_j\right) dx$$

$$\leq \int_C \inf_{z\geq 1}\left(z^{n-\bar{k}} \prod_{\{i,j\}\in M_0^c} \left(\bar{x}_i\bar{x}_j + zx_i\bar{x}_j + z\bar{x}_ix_j\right)\right) dx,$$

(z^m increases with m, if $z > 1$). Now,

$$\bar{x}_i\bar{x}_j + zx_i\bar{x}_j + z\bar{x}_ix_j = 1 + (1-2z)x_ix_j + (z-1)(x_i + x_j),$$

so ("exponentiating" and using $\Sigma_i x_i^2 \leq s$)

$$z^{n-\bar{k}} \prod_{\{i,j\}\in M_0^c} (\cdot) \leq \exp\left[\frac{1-2z}{2}s^2 + s(2z-1+n-1(z-1)) + (n-\bar{k})\log z\right].$$

Applying then Lemma 3, we get

$$P(\exists\, \mathscr{M} \text{ s.t. } R(\mathscr{M}) \geq \bar{k}) \leq \frac{(n-1)!!}{(n-1)!}\int_0^\infty \inf_{z\geq 1}(\exp(\mathscr{H}(z,s)))\, ds,$$

where

$$\mathscr{H}(z,s) = \frac{1-2z}{2}s^2 + s(2z-1+n-1(z-1)) + (n-\bar{k})\log z + (n-1)\log s.$$

As a function of z, $\mathscr{H}$ is concave up, and has its absolute *minimum* at

$$\bar{z} = \frac{\bar{k}-n}{s(n+1)-s^2}. \tag{4.6}$$

Let s_1 be the *smaller* root of the quadratic equation

$$s^2 - s(n+1) + \bar{k} - n = 0;$$

a simple computation shows that

$$s_1 = \frac{\bar{k}}{n} + O(1) = (1+\varepsilon)n^{1/2} + O(1).$$

Define a function $z(s)$ as follows:

$$z(s) = \begin{cases} \bar{z}(s), & \text{if } s \leq s_1, \\ 1, & \text{if } s > s_1. \end{cases}$$

Clearly, $z(s) \geq 1$ for all s. So, we have a bound

$$P(\exists\, \mathscr{M} \text{ s.t. } R(\mathscr{M}) \geq \bar{k}) \leq \frac{(n-1)!!}{(n-1)!}\int_0^n \exp(h(s))\, ds, \tag{4.7}$$

$$h(s) =_{\text{def}} \mathscr{H}(z(s),s).$$

(a) $s \geq s_1$. Here

$$h(s) = -\tfrac{1}{2}s^2 + s + (n-1)\log s$$

and

$$\max_{s \geq s_1} h(s) = h(s_1) = \frac{n}{2}\log\left(\frac{n}{e}\right) - nf(\varepsilon) + O(n^{1/2}). \tag{4.8}$$

So, the contribution of these values of s to the bound in (4.7) is of order $O(\exp(-nf(\varepsilon) + O(n^{1/2}))$.

(b) $s \leq s_1$. Here, using (4.6),

$$h(s) = \tilde{h}(s) =_{\text{def}} \frac{s^2}{2} - (ns) + (\bar{k} - n)$$
$$+(n - \bar{k})\log\frac{\bar{k} - n}{s(n+1) - s^2} + (n-1)\log s,$$

and it can be checked that $\tilde{h}(s)$ is concave down for $s \leq n^{\sigma}$, $\forall\ \sigma < 3/4$, and, furthermore,

$$\max_{s \leq n^{\sigma}} \tilde{h}(s) = \tilde{h}(s_2), \qquad s_2 = \frac{\bar{k}}{n} + O(1).$$

Since $\tilde{h}''(s) = O(n^{1/2})$ for s of order $n^{1/2}$, we obtain then [via Taylor's approximation of $\tilde{h}(s_1)$ $(= h(s_1))$ at s_2]:

$$\begin{aligned}\tilde{h}(s_2) &= h(s_1) - \tilde{h}''(\hat{s})(s_1 - s_2)^2/2 \quad (\hat{s} \text{ lies between } s_1 \text{ and } s_2)\\ &= h(s_1) + O(n^{1/2}).\end{aligned}$$

Invoking (4.8), we obtain that

$$\frac{(n-1)!!}{(n-1)!}\int_0^{s_1} \exp(h(s))\, ds = O\big(\exp(-nf(\varepsilon) + O(n^{1/2}))\big),$$

too.

The proof of Theorem 4 is now complete. □

An interesting consequence of Theorem 3 is that for a "typical" instance of the problem, we can greatly reduce all the preference lists, leaving in each member's list only the corresponding $cn^{1/2}\log n$ top choices, and declaring the remaining $(n - 1 - cn^{1/2}\log n)$ options unacceptable to this member. If $c > 1$, then almost surely no stable matching will be lost.

APPENDIX

PROPOSITION. *Let $f(2\nu, \mu)$ denote the total number of ways to arrange a 2ν-element set in μ disjoint circuits, of even length each. Then*

$$\lim_{\nu \to \infty} ((2\nu)!)^{-1} \sum_{\mu} 2^{2\mu} f(2\nu, \mu) = e^{-1}.$$

PROOF. Since μ circuits can be oriented in 2^{μ} ways, $g(2\nu, \mu) =_{\text{def}} 2^{\mu} f(2\nu, \mu)$ is the total number of permutations of the set $\{1, 2, \ldots, 2\nu\}$ with μ cycles, each of even length greater than or equal to 4. A general enumerative identity for permutations with restricted cycle lengths (e.g., Sachkov [25]) implies that

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\sum_{l=0}^{k} g(k, l) x^l \right) = \exp(x a(t)),$$

where

$$a(t) = \sum_{j=2}^{\infty} \frac{t^{2j}}{2j} = \frac{1}{2} \left(\log \frac{1}{1 - t^2} - t^2 \right).$$

Therefore,

$$\sum_{k=0}^{\infty} \frac{t^k}{k!} \left(\sum_{l=0}^{k} g(k, l) 2^l \right) = \frac{e^{-t^2}}{1 - t^2},$$

and the statement follows via a standard application of the Darboux formula (Bender [1]).

Acknowledgments. Like [22], this paper is an outgrowth of the work started during my visit to the Computer Science Department, Stanford University. It is my pleasure to thank Don Knuth for hospitality, and for generously sharing with me his knowledge of the stable matching problem and its probabilistic angle. The joint work with Don and Rajeev Motwani [18] helped me greatly to appreciate the algorithmic nature of the problem. I am very grateful to Tomás Feder and Rob Irving for stimulating discussions of possible directions of further research. The results of this paper and/or [22] have been presented in several seminars, and I want to express my gratitude for these opportunities to Fan Chung, Peter Winkler, Michal Karoński (Bellcore), Alan Frieze (Carnegie Mellon University), Zevi Miller (Miami University), Wojbor Woyczyński (Case Western Reserve University), Mike Fligner, Joe Verducci (AMS-IMS-SIAM Conference, Amherst) and Ricky Vohra (ORSA Conference, Monterey). My sincere thanks go to a referee for many helpful suggestions, especially for bringing to my attention an embarrassing gap in the original proof of Lemma 2.

NOTE ADDED IN PROOF. Very recently, Rob Irving and I proved that $\limsup P(S_n \geq 1) \leq e^{1/2}/2$ in "An upper bound for the solovability probability of a random stable roommates instance" submitted to *Combinatorics, Probability and Computing Journal*.

REFERENCES

[1] BENDER, E. A. (1974). Asymptotic methods in enumeration. *SIAM Rev.* **16** 485–514.
[2] BREIMAN, L. (1968). *Probability*. Addison-Wesley, Reading, MA.
[3] 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.

[4] DARLING, D. A. (1953). On a class of problems related to the random division of an interval. *Ann. Math. Statist.* **24** 239–253.

[5] DE BRUIJN, N. G. (1958). *Asymptotic Methods in Analysis*. North-Holland, Amsterdam.

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

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

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

[9] GRAHAM, R. L., KNUTH, D. E. and PATASHNIK, O. (1989). *Concrete Mathematics*. Addison-Wesley, Reading, MA.

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

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

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

[13] KARLIN, S. (1981). Some natural viability systems for a multiallelic locus: a theoretical study. *Genetics* **97** 457–473.

[14] KINGMAN, J. F. C. (1988). Typical polymorphisms maintained by selection at a single locus. *J. Appl. Probab.* **25** 113–125.

[15] KINGMAN, J. F. C. (1989). Maxima of random quadratic forms on a simplex. In *Probability, Statistics and Mathematics: Papers in Honor of Samuel Karlin* (T. W. Anderson, K. B. Athreya and D. L. Inglehart, eds.). 123–140.

[16] KNUTH, D. E. (1976). *Marriages Stables et leurs relations avec d'autres problémes combinatoires*. Les Presses de l'Université de Montréal.

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

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

[19] LÉVY, P. (1939). Sur la division d'un segment par des points choisis au hasard. *Comptes Rendus Acad. Sci. Paris* **208** 147–149.

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

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

[22] PITTEL, B. (1992). On likely solutions of a stable marriage problem. *Ann. Appl. Probab.* **2** 358–401.

[23] RÉNYI, A. (1970). *Probability Theory*. North-Holland, New York.

[24] SACHKOV, V. N. (1978). *Probabilistic Methods in Combinatorial Analysis*. Nauka, Moscow Amsterdam (in Russian).

[25] SWAMY, M. N. S. and THULASIRAMAN, K. (1981). *Graphs, Networks and Algorithms*. Wiley, New York.

DEPARTMENT OF MATHEMATICS
OHIO STATE UNIVERSITY
COLUMBUS, OHIO 43210