Asymptotic enumeration of sparse 0-1 matrices with irregular row and column sums

Catherine Greenhill
School of Mathematics
The University of New South Wales
Sydney NSW 2052, Australia
csg@unsw.edu.au

Brendan D. McKay
Department of Computer Science
Australian National University
Canberra, ACT 0200, Australia
bdm@cs.anu.edu.au

Xiaoji Wang
Maritime Operations Division
DSTO, PO Box 1500
Edinburgh, SA 5111, Australia
Jimmy.Wang@defence.gov.au

17th March 2005

Abstract

Let $\boldsymbol{s} = (s_1, \ldots, s_m)$ and $\boldsymbol{t} = (t_1, \ldots, t_n)$ be vectors of non-negative integer-valued functions with equal sum $S = \sum_{i=1}^m s_i = \sum_{j=1}^n t_j$. Let $N(\boldsymbol{s}, \boldsymbol{t})$ be the number of $m \times n$ matrices with entries from $\{0, 1\}$ such that the ith row has row sum s_i and the jth column has column sum t_j. Equivalently, $N(\boldsymbol{s}, \boldsymbol{t})$ is the number of labelled bipartite graphs with degrees of the vertices in one side of the bipartition given by $\boldsymbol{s}$ and the degrees of the vertices in the other side given by $\boldsymbol{t}$. We give an asymptotic formula for $N(\boldsymbol{s}, \boldsymbol{t})$ which holds when $S \to \infty$ with $1 \le st = o(S^{2/3})$, where $s = \max_i s_i$ and $t = \max_j t_j$. This extends a result of McKay and Wang (2003) for the semiregular case (when $s_i = s$ for $1 \le i \le m$ and $t_j = t$ for $1 \le j \le n$). The previously strongest result for the non-semiregular case required $1 \le \max\{s, t\} = o(S^{1/4})$, due to McKay (1984).

1 Introduction

The problem of obtaining asymptotic formulae for the number of 0-1 matrices with given row and column sums (equivalently, the number of bipartite graphs with fixed degree sequences) has received much attention. The asymptotics are with respect to the number of 1s in the matrix; equivalently, the number of edges in the graph.

Let $\boldsymbol{s} = (s_1, s_2, \ldots, s_m)$ and $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ be sequences of nonnegative integers such that $\sum_{i=1}^m s_i = \sum_{j=1}^n t_j$. Define $\mathcal{M}(\boldsymbol{s}, \boldsymbol{t})$ to be the class of 0-1 matrices of order $m \times n$ such that the sum of row i is s_i and the sum of column j is t_j, for each i, j. Each $M \in \mathcal{M}(\boldsymbol{s}, \boldsymbol{t})$ corresponds to a simple bipartite graph $G(M)$, with vertices $X \cup Y$ where $X = \{x_1, x_2, \ldots, x_m\}$ and $Y = \{y_1, y_2, \ldots, y_n\}$ (assumed disjoint). Vertex x_i is adjacent to vertex y_j if and only if $(M)_{ij} = 1$. Also, vertex x_i has degree s_i and vertex y_j has degree t_j. Define $N(\boldsymbol{s}, \boldsymbol{t}) = |\mathcal{M}(\boldsymbol{s}, \boldsymbol{t})|$. Let S be defined by $S = \sum_{i=1}^m s_i = \sum_{j=1}^n t_j$. Also let $s = \max_i s_i$ and $t = \max_j t_j$. A matrix with equal row sums and equal column sums, and the corresponding graph, will be called *semiregular*.

Study of the asymptotic behaviour of $N(\boldsymbol{s}, \boldsymbol{t})$ began with Read [12], who solved the semiregular case for $s = t = 3$. The semiregular case for arbitrary but fixed s and t was solved by Everett and Stein [5]. Békéssy, Békéssy and Komlós [1], Bender [2], and Wormald [14] independently extended this to arbitrary row and column sums provided s and t are bounded.

The first attempt to allow s and t to grow with S was by O'Neil [11], who solved the semiregular case for $s, t \le (\log n)^{1/4-\epsilon}$. Improvements that still allowed at most fractional logarithmic growth of s and t were obtained by Mineev and Pavlov [10] and by Bollobás and McKay [3].

McKay [6] applied a completely different method (the ancestor of the method we will use here) to obtain the asymptotic value of $N(\boldsymbol{s}, \boldsymbol{t})$ whenever $\max\{s, t\} = o(S^{1/4})$.

For any x, define $[x]_0 = 1$ and, for integer $k > 0$, $[x]_k = x(x-1)\cdots(x-k+1)$. Also define $S_k = \sum_{i=1}^m [s_i]_k$ and $T_k = \sum_{j=1}^n [t_j]_k$ for $k \ge 1$. Note that $S_1 = T_1 = S$.

Theorem 1.1. [6] *Suppose that $S \to \infty$ and $1 \le \max\{s, t\}^2 < cS$ for some constant $c < \frac{1}{6}$. Then*

$$N(\boldsymbol{s}, \boldsymbol{t}) = \frac{S!}{\prod_{i=1}^m s_i! \prod_{j=1}^n t_j!} \exp\biggl(-\frac{S_2 T_2}{2S^2} + O\bigl(\max\{s, t\}^4 / S\bigr)\biggr).$$

Of course the error term in Theorem 1.1 is only $o(1)$ if $\max\{s, t\} = o(S^{1/4})$. That range was extended in the semiregular case by McKay and Wang [7].

Theorem 1.2. [7] *Suppose that $S \to \infty$ and $1 \le st = o(S^{2/3})$. In the semiregular case, $N(\boldsymbol{s}, \boldsymbol{t})$ is given by*

$$\frac{S!}{(s!)^m (t!)^n} \exp\biggl(-\frac{(s-1)(t-1)}{2} - \frac{(s-1)(t-1)(2st - s - t + 2)}{12S} + O\Bigl(\frac{s^3 t^3}{S^2}\Bigr)\biggr).$$

A different range of the same problem, when the density (S/mn) is high, has been solved by Canfield and McKay [4] using analytic methods.

Our aim in this paper is to extend both Theorems 1.1 and 1.2 to the non-semiregular case.

Theorem 1.3. *Suppose that $S \to \infty$, and that $\boldsymbol{s} = (s_1, \ldots, s_m)$, $\boldsymbol{t} = (t_1, \ldots, t_n)$ are vectors of nonnegative integer functions of S such that $\sum_i s_i = \sum_j t_j = S$. If $1 \le st = o(S^{2/3})$ then*

$$N(\boldsymbol{s},\boldsymbol{t}) = \frac{S!}{\prod_{i=1}^m s_i! \prod_{j=1}^n t_j!} \times \exp\biggl(-\frac{S_2T_2}{2S^2} - \frac{S_2T_2}{2S^3} + \frac{S_3T_3}{3S^3} - \frac{S_2T_2(S_2+T_2)}{4S^4} - \frac{{S_2}^2T_3 + S_3{T_2}^2}{2S^4} + \frac{{S_2}^2{T_2}^2}{2S^5} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\biggr).$$

A weaker version of the above theorem appeared in the Ph.D. thesis of the third author [13].

In the next section we describe the model used and outline our approach. This is essentially the same as in [7], but the lack of semiregularity causes many technical difficulties that were not present before.

A note on our usage of the $O(\,)$ notation in the following is in order since we use it very strictly. Given a fixed function $f(S) = o(S^{2/3})$, and any quantity ϕ that depends on any of our variables, $O(\phi)$ denotes any quantity whose absolute value is bounded above by $|c\phi|$ for some constant c that depends on f *and nothing else*, provided $1 \le st \le f(S)$. Note that this includes the case where $\phi = 0$.

2 The model and our approach

We use the same model as in [7], but for completeness we describe it again here. Our calculations are performed in the pairings model. Consider a set of S *points* arranged in *cells* $x_1, x_2, \ldots, x_m$, where cell x_i has size s_i for $1 \le i \le m$, and another set of S points arranged in cells $y_1, y_2, \ldots, y_n$ where cell y_j has size t_j for $1 \le j \le n$. Take a partition P (called a *pairing*) of the $2S$ points into S *pairs* with each pair having the form (x, y) where $x \in x_i$ and $y \in y_j$ for some i, j. A *random pairing* is such a pairing chosen uniformly at random. It contains S pairs.

Two pairs are called *parallel* if they involve the same cells. The *multiplicity* of a pair is the number of pairs (including itself) parallel to it. A *simple* pair is a pair of multiplicity one. A *double* pair is a set of two parallel pairs of multiplicity two, while a *triple* pair is a set of three parallel pairs of multiplicity three. If p is a point, then $v(p)$ is the cell containing that point.

The first lemma is easy.

Lemma 2.1. *For $0 \le r \le S$, the probability of r given pairs occurring in a random pairing is $1/[S]_r$.*

Define $P(\boldsymbol{s},\boldsymbol{t})$ to be the probability that P contains no pairs of multiplicity greater than one. Since each matrix in $\mathcal{M}(\boldsymbol{s},\boldsymbol{t})$ corresponds to exactly $\prod_{i=1}^m s_i! \prod_{j=1}^n t_j!$ such pairings, we have

$$N(\boldsymbol{s},\boldsymbol{t}) = \frac{S!}{\prod_{i=1}^m s_i! \prod_{j=1}^n t_j!} P(\boldsymbol{s},\boldsymbol{t}). \tag{2.1}$$

Our task is thus reduced to computing $P(\boldsymbol{s},\boldsymbol{t})$.

We begin with some cases where the expected number of pairs of multiplicity greater than one is quite small, since removal of these cases from our main proof will lead to some welcome simplications. Say that the pair (S_2, T_2) is *substantial* if the following conditions hold:

- $1 \le st = o(S^{2/3})$,
- $S_2 \ge s \log^2 S$ and $T_2 \ge t \log^2 S$,
- $S_2 T_2 \ge (st)^{3/2} S$.

When (S_2, T_2) is not substantial, we can prove Theorem 1.3 using inclusion-exclusion. Throughout this paper we often use the fact that $S_r \le sS_{r-1}$ and $T_r \le tT_{r-1}$ for any $r \ge 2$. In the following lemma we also use the fact that $S_{2r} \le {S_2}^r$ and $T_{2r} \le {T_2}^r$ for $r \ge 1$.

Lemma 2.2. *If $1 \le st = o(S^{2/3})$ but (S_2, T_2) is not substantial, then the conclusion of Theorem 1.3 holds.*

Proof. Take a random pairing. An *error* is an unordered set of 2 parallel pairs. We use inclusion-exclusion to estimate the probability $P(\boldsymbol{s},\boldsymbol{t})$ that there are no errors.

Using Lemma 2.1, we find that the total of the probabilities of all possible sets of 4 distinct errors is $O(s^3t^3/S^2)$. For example, the cases where 1 cell of X and 4 cells of Y are involved contributes

$$O\left(\frac{S_8 {T_2}^4}{S^8}\right) = O\left(\frac{{S_2}^4 {T_2}^4}{S^8}\right),$$

which is easily seen to be $O(s^3t^3/S^2)$ if (S_2, T_2) is not substantial. Similarly we can see that the contribution from each case of 4 distinct errors is $O(s^3t^3/S^2)$, and there are only a finite number of cases.

By the Bonferroni inequalities, we can thus restrict ourselves to sets of 3 or fewer errors. Furthermore, the only arrangement of 3 distinct errors which can contribute more than $O(s^3t^3/S^2)$ is when the 3 errors consist of each subset of two pairs from a triple of parallel pairs. (This can be seen by checking cases.)

The total probability for all placements of 1 error is $S_2T_2/2[S]_2$. Similarly, the case where 2 errors have a pair in common has a total probability of $S_3T_3/2[S]_3$. The case of 3 errors which involve only 3 parallel pairs altogether gives a total probability of $S_3T_3/6[S]_3$.

The remaining contributing situation is for 2 errors which do not have any pairs in common. Suppose the errors are $\{(p_1, p_1'), (p_2, p_2')\}$ and $\{(p_3, p_3'), (p_4, p_4')\}$, where $v(p_1) = v(p_2) = i$, $v(p_1') = v(p_2') = i'$, $v(p_3) = v(p_4) = j$, and $v(p_3') = v(p_4') = j'$. We count the number of ordered 8-tuples $(p_1, p_2, p_3, p_4, p_1', p_2', p_3', p_4')$ by summing the expression which holds for $i \ne j$ and $i' \ne j'$, then separately correcting the cases $i = j$ and $i' = j'$, then finally correcting the case where both $i = j$ and $i' = j'$ together. This gives

$$\begin{aligned}
&\sum_{i,j=1}^{m} \sum_{i',j'=1}^{n} [s_i]_2[s_j]_2\, [t_{i'}]_2[t_{j'}]_2 \\
&\qquad + \sum_{i=1}^{m} \sum_{i',j'=1}^{n} \bigl([s_i]_4 - [s_i]_2{}^2\bigr)\, [t_{i'}]_2[t_{j'}]_2 + \sum_{i,j=1}^{m} \sum_{i'=1}^{n} [s_i]_2[s_j]_2\, \bigl([t_{i'}]_4 - [t_{i'}]_2{}^2\bigr) \\
&\qquad + \sum_{i=1}^{m} \sum_{i'=1}^{n} \bigl([s_i]_4[t_{i'}]_4 + [s_i]_2{}^2[t_{i'}]_2{}^2 - [s_i]_4[t_{i'}]_2{}^2 - [s_i]_2{}^2[t_{i'}]_4\bigr) \\
&\quad = (S_2{}^2 - 4S_3 - 2S_2)(T_2{}^2 - 4T_3 - 2T_2).
\end{aligned}$$

This uses the fact that $[x]_2{}^2 = [x]_4 + 4[x]_3 + 2[x]_2$. Each pair of 2 errors of this type corresponds to 8 such 8-tuples, so the total probability in this case is

$$\frac{(S_2{}^2 - 4S_3 - 2S_2)(T_2{}^2 - 4T_3 - 2T_2)}{8[S]_4}.$$

Combining these contributions using the inclusion-exclusion formula, we find that

$$P(\boldsymbol{s}, \boldsymbol{t}) = 1 - \frac{S_2T_2}{2[S]_2} + \frac{S_3T_3}{2[S]_3} + \frac{(S_2{}^2 - 4S_3 - 2S_2)(T_2{}^2 - 4T_3 - 2T_2)}{8[S]_4} - \frac{S_3T_3}{6[S]_3} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr).$$

After multiplication by $S!/(\prod_{i=1}^m s_i! \prod_{j=1}^n t_j!)$ this is equal to the expression in Theorem 1.3, under our present assumptions. (Note that since (S_2, T_2) is not substantial, the term $S_2{}^2T_2{}^2/2S^5$ which appears in the statement of Theorem 1.3 is absorbed into the error term.) □

In view of Lemma 2.2, we can assume that (S_2, T_2) is substantial from now on. Our next task will be to bound the number of double and triple pairs, and show that pairs of higher multiplicity make asymptotically insignificant contribution. Define

$$\begin{aligned}
N_2 &= \begin{cases} 8 & \text{if } S_2T_2 < S^{7/4}, \\ \lceil \log(S) \rceil & \text{if } S^{7/4} \le S_2T_2 < S^2 \log S/21, \\ \lceil 21S_2T_2/S^2 \rceil & \text{if } S^2 \log S/21 \le S_2T_2; \end{cases} \\
N_3 &= \max\bigl(\lceil \log(S) \rceil, \lceil 7S_3T_3/S^3 \rceil\bigr).
\end{aligned}$$

For $d, h \ge 0$, define $\mathcal{C}_{d,h} = \mathcal{C}_{d,h}(\boldsymbol{s}, \boldsymbol{t})$ to be the set of all pairings with exactly d double pairs and h triple pairs, but no pairs of multiplicity greater than 3. With high probability, a random pairing has no more than N_2 double pairs and no more than N_3 triple pairs. In fact we can prove the following.

Lemma 2.3. *If (S_2, T_2) is substantial then*

$$\frac{1}{P(\boldsymbol{s}, \boldsymbol{t})} = \bigl(1 + O(s^3t^3/S^2)\bigr) \sum_{d=0}^{N_2} \sum_{h=0}^{N_3} \frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{0,0}|}.$$

Proof. Let P be a random pairing. Define P_1 to be the probability that P contains a pair of multiplicity greater than three, which is at most equal to the expectation of the number of sets of 4 parallel pairs. By Lemma 2.1, we have

$$P_1 \leq \tfrac{1}{24} S_4 T_4 / [S]_4 = O(S_4 T_4 / S^4) = O(s^3 t^3 / S^2).$$

Let $d = N_2 + 1$ and define P_2 to be the probability that P has at least d double pairs, which is at most equal to the expectation of the number of sets of d double pairs. By Lemma 2.1, we have

$$P_2 \leq \binom{S_2 T_2/2}{d} / [S]_{2d} \leq \binom{S_2 T_2/2}{d} (S - 2d)^{-2d}.$$

In the case that $S_2 T_2 < S^{7/4}$ we have that $d = 9$ and $P_2 = O(S^{-2})$. In the other cases we have that both $d > \log S$ and $d > 21 S_2 T_2 / S^2$ and so

$$\begin{aligned} P_2 &\leq \left(\frac{S_2 T_2 e}{2d(S - 2d)^2} \right)^d, \quad \text{since } d! \geq (d/e)^d \\ &\leq \left(\frac{e(1 + o(1))}{42} \right)^d, \quad \text{since } d = o(S) \text{ and } d > 21 S_2 T_2 / S^2 \\ &= O(S^{-2}), \quad \text{since } d > \log S \text{ and } \log(e/42) < -2. \end{aligned}$$

By the same argument, the probability P_3 that P has at least $N_3 + 1$ triple pairs is $O(S^{-2})$.

Let A be the set of all pairings and $B \subseteq A$ be the set of all the pairings which have a pair of multiplicity greater than 3, or have more than N_2 double pairs, or have more than N_3 triple pairs. Then we have

$$P(\boldsymbol{s}, \boldsymbol{t}) = \frac{|\mathcal{C}_{0,0}|}{|A|}.$$

Hence,

$$\begin{aligned} \frac{1}{P(\boldsymbol{s}, \boldsymbol{t})} &= \frac{|A| - |B|}{|\mathcal{C}_{0,0}|} \left(\frac{|A|}{|A| - |B|} \right) \\ &= \sum_{d=0}^{N_2} \sum_{h=0}^{N_3} \frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{0,0}|} \left(1 + \frac{|B|/|A|}{1 - |B|/|A|} \right) \\ &= \sum_{d=0}^{N_2} \sum_{h=0}^{N_3} \frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{0,0}|} \big(1 + O(s^3 t^3 / S^2) \big), \end{aligned}$$

since $|B|/|A| \leq P_1 + P_2 + P_3$. □

Now our task is reduced to calculating the ratios $|\mathcal{C}_{d,h}| \,/\, |\mathcal{C}_{0,0}|$ for $0 \leq d \leq N_2$ and $0 \leq h \leq N_3$, in the case that (S_2, T_2) is substantial. We do this by extending the argument given by the second and third authors in [7] for the semiregular case: namely, by applying operations on pairings called *switchings*.

We will make use of the following two operations on pairings.

d-switching:

Take a double pair $\{(p_1, p_1'), (p_2, p_2')\}$ and two simple pairs (p_3, p_3') and (p_4, p_4'), such that six distinct cells are involved. Replace these four pairs by (p_1, p_3'), (p_2, p_4'), (p_3, p_1') and (p_4, p_2'), which must be simple.

t-switching:

Take a triple pair $\{(p_1, p_1'),\ (p_2, p_2'),\ (p_3, p_3')\}$ and three simple pairs (p_4, p_4'), (p_5, p_5') and (p_6, p_6'), such that eight distinct cells are involved. Replace these six pairs by (p_1, p_4'), (p_2, p_5'), (p_3, p_6'), (p_4, p_1'), (p_5, p_2'), and (p_6, p_3'), which must be simple.

In Figure 1, which illustrates the two types of switchings, the cells are indicated by shaded ellipses and the pairs are indicated by line segments.

The inverse of a d-switching is called an *inverse d-switching*, and similarly for t-switchings.

Note that a t-switching reduces the number of triple pairs by one without affecting the number of double pairs, while a d-switching reduces the number of double pairs by one, without affecting the number of triple pairs. This allows us to estimate the ratios $|\mathcal{C}_{d,h}| \,/\, |\mathcal{C}_{d,h-1}|$ and $|\mathcal{C}_{d,0}| \,/\, |\mathcal{C}_{d-1,0}|$, respectively, which are then combined to give the required ratios $|\mathcal{C}_{d,h}| \,/\, |\mathcal{C}_{0,0}|$. These arguments are given in Section 4. First we must obtain fairly precise asymptotic estimates for certain quantities which will be needed. These calculations are given in the next section.

3 Random pairings

Throughout this section, P is a random pairing. Note that P contains S pairs. For later convenience, we note a few consequences of the definition of N_2 for substantial (S_2, T_2).

Lemma 3.1.
Suppose that (S_2, T_2) is substantial and that $0 \le d \le N_2$. Then

$$d^{3/2}s = o(S_2),\ \ d^{3/2}t = o(T_2)\ \ \textit{and}\ \ d^4 + d^3st + d^2s^2t^2 + s^3t^3 = o(S_2T_2).$$

Proof. Note that $d \le \lceil \log S \rceil$ or $d = O(S_2T_2/S^2)$. (Of course both may be true.) To prove the first claim, if $d \le \lceil \log S \rceil$ then

$$d^{3/2}s = O(s \log^{3/2} S) = O(S_2/\log^{1/2} S) = o(S_2).$$

If $d = O(S_2T_2/S^2)$, then

$$d^3s^2/{S_2}^2 = O(S_2{T_2}^3s^2/S^6) = O(s^3t^3/S^2) = o(1).$$

The second claim is proved analogously.

Next, note that

$$s^3t^3 \le (st)^{3/2}S_2T_2/S = o(S_2T_2).$$

If $d = O(S_2T_2/S)$ then

$$d^4 + d^3st + d^2s^2t^2 = O\big((S_2T_2)^4/S^8 + st(S_2T_2)^3/S^6 + s^2t^2(S_2T_2)^2/S^4\big) = o(S_2T_2).$$

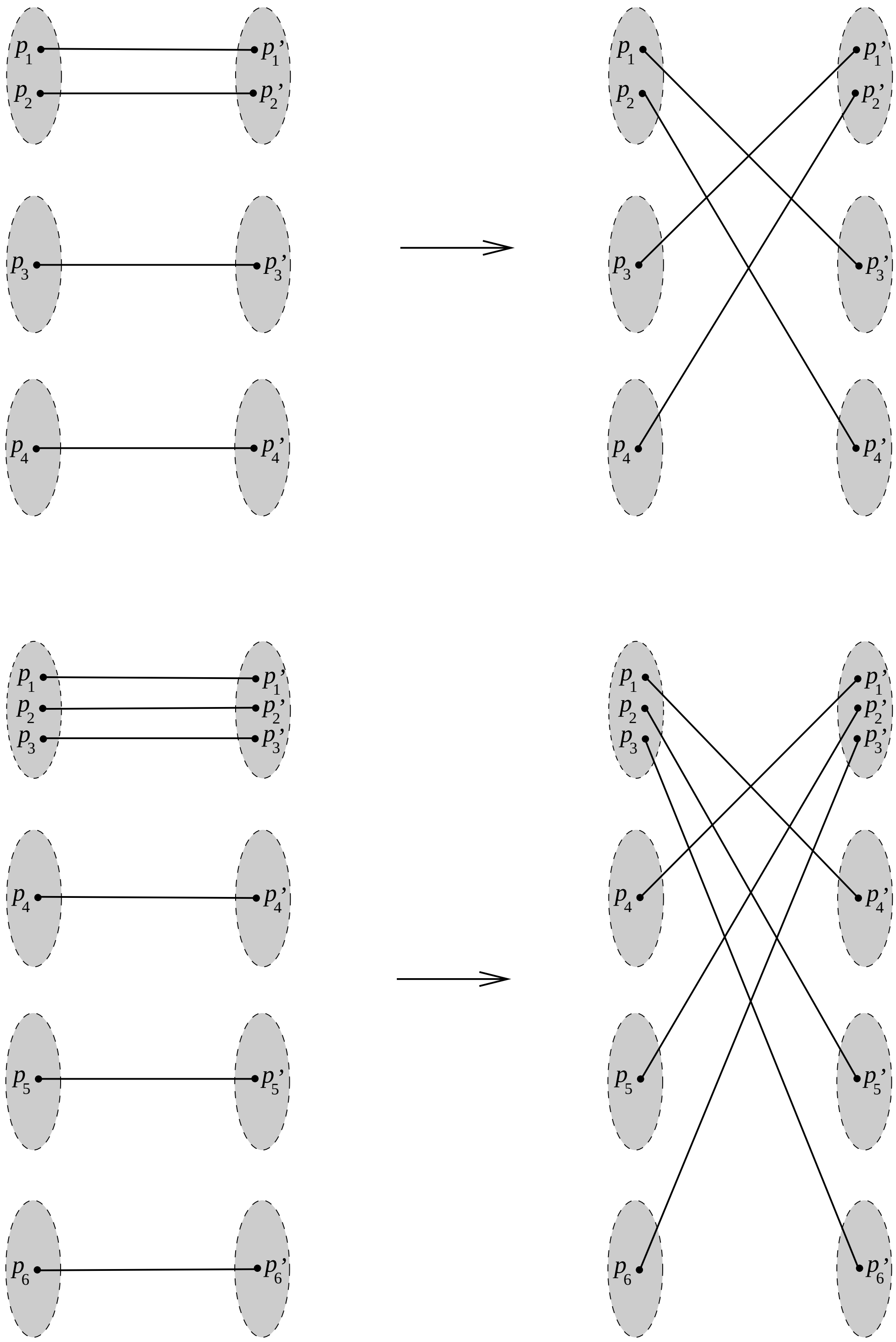

Figure 1: A d-switching (top) and a t-switching (bottom)

If $d = \lceil \log S \rceil$ then

$$\begin{aligned}(d^4 + d^3 st + d^2 s^2 t^2)/S_2 T_2 &= O\bigl(\log^4 S + st \log^3 S + s^2 t^2 \log^2 S\bigr)/S_2 T_2 \\ &= O\biggl(\frac{\log^4 S}{(st)^{3/2} S} + \frac{\log^3 S}{(st)^{1/2} S} + \frac{(st)^{1/2} \log^2 S}{S}\biggr) \\ &= o(1).\end{aligned}$$ □

If K is a bipartite multigraph, let $e(K)$ denote its number of edges (counting multiplicities). If xx' is an edge of K, then $\mu_K(xx')$ denotes the multiplicity of the edge between x and x', or 0 if there is no such edge. If K and K' are bipartite multigraphs with the same vertex set, then $K + K'$ is the bipartite multigraph with the same vertex set such that $\mu_{K+K'}(xx') = \mu_K(xx') + \mu_{K'}(xx')$ for all (x, x'). Similarly, $2K$ means $K + K$ and $K + xx'$ is the same as K except that $\mu_{K+xx'}(xx') = \mu_K(xx') + 1$.

Let L be a simple bipartite graph with parts X and Y, and let H be a bipartite multigraph on the same vertex set with the restriction that if any edge xx' has $\mu_H(xx') \geq 1$, then xx' is an edge of L. Let ℓ and ℓ' denote the maximum degrees of L in the X part and the Y part, respectively.

Given a pairing P, the *bipartite multigraph $B(P)$ associated with P* has parts X and Y. The edges of $B(P)$ are in correspondence with the pairs of P: the pair (x, y) corresponds to an edge $\{v(x), v(y)\}$.

Define $\mathcal{C}(L, H) = \mathcal{C}(L, H; \boldsymbol{s}, \boldsymbol{t})$ to be the set of all pairings P such that the following are true for all (x, x'):

- If xx' is an edge of L, then $\mu_{B(P)}(xx') = \mu_H(xx')$.
- If xx' is not an edge of L, then $\mu_{B(P)}(xx') \leq 1$.

In other words, $B(P)$ must be simple outside L and match H inside L.

Lemma 3.2. *Suppose that L is as defined above, and that H and $H + K$ satisfy the requirements given above for H. Let h_i, h'_j be the degrees of x_i, y_j in H, respectively, and similarly k_i, k'_j for K ($1 \leq i \leq m$, $1 \leq j \leq n$). Then, if $(st + s\ell' + \ell t)e(K) = o(S)$, $e(H) = o(S)$, and $\mathcal{C}(L, H) \neq \emptyset$, we have*

$$\begin{aligned}&\frac{|\mathcal{C}(L, H+K)|}{|\mathcal{C}(L, H)|} \\ &\quad = \frac{\prod_{i=1}^m [s_i - h_i]_{k_i} \prod_{j=1}^n [t_j - h'_j]_{k'_j}}{[S - e(H)]_{e(K)} \prod_{(x,x') \in X \times Y} [\mu_{H+K}(xx')]_{\mu_K(xx')}} \bigl(1 + O((st + s\ell' + \ell t)e(K)/S)\bigr).\end{aligned}$$

Proof. Apart from the form of the error term, this is a special case of the combination of Theorems 3.4 and 3.8 of [6] (but note that the inequality in Theorem 3.8 was printed with "≤" when it is really "≥"). The error term in [6] is written in terms of $\max\{s, t\}$ and $\max\{\ell, \ell'\}$, but careful inspection of the proof (especially [6, Lemmas 3.2 and 3.6]) shows that the error term we give here is established. □

We will use Lemma 3.2 to analyse the structure of $\mathcal{C}_{d,0}$. For a pairing $P \in \mathcal{C}_{d,0}$, let $D(P)$ be the simple bipartite graph with parts X and Y and just those edges which correspond in position to the d double pairs of P. The next lemma is [7, Lemma 4].

Lemma 3.3. *Let $D = D(P')$ for some $P' \in \mathcal{C}_{d,0}$, where $0 \le d \le N_2$. Let A be a simple bipartite graph with parts X and Y which is edge-disjoint from D. Let d_i, d'_j be the degrees of x_i, y_j in D, respectively, and define a_i, a'_j similarly for A. Suppose that $e(A) = o(S/st)$. Then the probability that $A \subseteq B(P)$, when P is chosen at random from those $P \in \mathcal{C}_{d,0}$ such that $D(P) = D$, is*

$$\frac{\prod_{i=1}^{m}[s_i - 2d_i]_{a_i} \prod_{j=1}^{n}[t_j - 2d'_j]_{a'_j}}{[S - 2d]_{e(A)}} \big(1 + O(st/S)\, e(A)\big).$$

In the next lemma we prove two useful, easy results. The latter involves the functions $f_k = f_k(D)$ and $f'_k = f'_k(D)$ defined on $X \times X$ and $Y \times Y$ as follows. Let $k \in \mathbb{Z}^+$ and let D be a simple bipartite graph with parts X and Y representing the position of double pairs of some pairing. Let d_i, d'_j be the degrees of x_i, y_j in D, respectively, for all $x_i \in X$, $y_j \in Y$. Then define

$$f_k(v_1, v_2) = \begin{cases} [s_{v_1} - 2d_{v_1}]_k\, [s_{v_2} - 2d_{v_2}]_2 & \text{if } v_1 \neq v_2, \\ [s_{v_1} - 2d_{v_1}]_{k+2} & \text{if } v_1 = v_2, \end{cases}$$

$$f'_k(w_1, w_2) = \begin{cases} [t_{w_1} - 2d'_{w_1}]_k\, [t_{w_2} - 2d'_{w_2}]_2 & \text{if } w_1 \neq w_2, \\ [t_{w_1} - 2d'_{w_1}]_{k+2} & \text{if } w_1 = w_2. \end{cases}$$

Lemma 3.4.
(i) For any constant $r \ge 1$ we have $\sum_{i=1}^{m}[s_i - 2d_i]_r = S_r + O(ds^{r-1})$ and $\sum_{j=1}^{n}[t_j - 2d'_j]_r = T_r + O(dt^{r-1})$.
(ii) Suppose that (S_2, T_2) is substantial. For $k \in \{1, 2\}$ we have

$$\sum_{v_1, v_2 \in X} f_k(v_1, v_2) = S_k S_2 \left(1 + O\left(\frac{ds}{S_2} + \frac{s^{k-1}}{S_k}\right)\right)$$

and

$$\sum_{w_1, w_2 \in Y} f'_k(w_1, w_2) = T_k T_2 \left(1 + O\left(\frac{dt}{T_2} + \frac{t^{k-1}}{T_k}\right)\right).$$

Proof. Consider the first statement of (i). For each i, we have

$$[s_i - 2d_i]_r = [s_i]_r + d_i\, p(d_i, s_i)$$

where $p(x, y)$ is a polynomial of total degree $r - 1$. Since $0 \le d_i \le s_i/2$, it follows that $p(d_i, s_i) = O(s^{r-1})$. Summing over i proves the first statement of (i), and the proof of the second statement is entirely analogous.

Now suppose that (S_2, T_2) is substantial. For a given $v_1 \in X$ we have

$$
\begin{aligned}
&\sum_{v_2\in X} f_k(v_1,v_2)\\
&= \Bigl(\sum_{v_2\in X} [s_{v_1}-2d_{v_1}]_k\,[s_{v_2}-2d_{v_2}]_2\Bigr) - \bigl([s_{v_1}-2d_{v_1}]_{k+2} - [s_{v_1}-2d_{v_1}]_k\,[s_{v_1}-2d_{v_1}]_2\bigr)\\
&= [s_{v_1}-2d_{v_1}]_k\,S_2\bigl(1+O(ds/S_2)\bigr) + \begin{cases} 2[s_{v_1}-2d_{v_1}]_2 & \text{if } k=1,\\ 4[s_{v_1}-2d_{v_1}]_2(s_{v_1}-2d_{v_1}-\frac{3}{2}) & \text{if } k=2\end{cases}\\
&= [s_{v_1}-2d_{v_1}]_k\,S_2\bigl(1+O(ds/S_2)\bigr) + O(s^{k-1}S_2)\\
&= [s_{v_1}-2d_{v_1}]_k\,S_2\bigl(1+O(ds/S_2+s^{k-1}/S_k)\bigr),
\end{aligned}
$$

using (i). Therefore

$$
\begin{aligned}
\sum_{v_1,v_2\in X} f_k(v_1,v_2) &= \sum_{v_1\in X}[s_{v_1}-2d_{v_1}]_k\,S_2\bigl(1+O(ds/S_2+s^{k-1}/S_k)\bigr)\\
&= S_kS_2\bigl(1+O(ds/S_2+s^{k-1}/S_k)\bigr)
\end{aligned}
$$

using (i) again. This proves the first statement in (ii), and the proof of the second statement is entirely analogous. □

Using Lemma 3.4(i) we can prove the following.

Lemma 3.5. *Let H be a fixed simple bipartite graph with vertices $\bar{X}\cup\bar{Y}$. Let h_v, h'_w denote the degree of vertices $v\in\bar{X}$, $w\in\bar{Y}$ in H and let $e(H)$ be the number of edges of H. Assume that the minimum degree of H is at least 1. Let k be the maximum degree in H over all vertices in $\bar{X}$, and let ℓ be the maximum degree in H over all vertices in $\bar{Y}$. Assume that $ds^{k-1}+s^k = o(S_k)$ and $dt^{\ell-1}+t^\ell = o(T_\ell)$. Then the expected number of injections ϕ from $\bar{X}$ into X and from $\bar{Y}$ into Y such that ϕ maps the edges of H onto simple edges of $B(P)$ is*

$$
S^{-e(H)}\prod_{i\in\bar{X}} S_{h_i}\prod_{j\in\bar{Y}} T_{h'_j}\ \left(1+O\biggl(\frac{st}{S}+\frac{ds^{k-1}+s^k}{S_k}+\frac{dt^{\ell-1}+t^\ell}{T_\ell}+\frac{ds^kt^\ell}{S_k\,T_\ell}\biggr)\right),
$$

where the error term assumes that H is fixed. (Note that the final error term might not be $o(1)$ under our assumptions.)

Proof. Let $\bar{X}=\{x_1,\ldots,x_p\}$ and $\bar{Y}=\{y_1,\ldots,y_q\}$. Fix $D=D(P')$ for some $P'\in\mathcal{C}_{d,0}$. Now choose a random $P\in\mathcal{C}_{d,0}$ such that $D(P)=D$. (We will find that our required expectation is independent of D, to within the required accuracy.) For some injection ϕ from $\bar{X}$ into X and $\bar{Y}$ into Y, let $v_i=\phi(x_i)$, $w_j=\phi(y_j)$ for $1\le i\le p$, $1\le j\le q$. Then the probability that ϕ maps each edge of H onto a simple pair of P is

$$
\frac{\prod_{i=1}^p[s_{v_i}-2d_{v_i}]_{h_i}\,\prod_{j=1}^q[t_{w_j}-2d'_{w_j}]_{h'_j}}{[S-2d]_{e(H)}}\bigl(1+O(st/S)\bigr),
$$

by Lemma 3.3, under the assumption that none of the edges of $\phi(H)$ belong to D. We need to sum this over all possible injections ϕ which do not map an edge of H onto a pair in D. First we sum over all injections without regard to the latter condition. We can achieve this by summing first over all v_1, and then over all $v_2 \neq v_1$, and so on. This gives

$$\begin{aligned}\sum_{\substack{(v_1,\ldots,v_p)\\ \text{distinct}}} \prod_{i=1}^{p} [s_{v_i} - 2d_{v_i}]_{h_i} &\\ &= \bigl(S_{h_1} - O(ds^{h_1-1})\bigr)\bigl(S_{h_2} - O(ds^{h_2-1} + s^{h_2})\bigr) \cdots \bigl(S_{h_p} - O(ds^{h_p-1} + s^p)\bigr)\\ &= \prod_{i=1}^{p}\bigl(S_{h_i} - O(ds^{h_i-1} + s^{h_i})\bigr),\end{aligned}$$

using Lemma 3.4(i). A similar result holds for the choices of q-tuples in Y with distinct entries.

We bound the relative contribution to this sum from those injections ϕ which map an edge of H onto an edge of D, as follows. Fix an edge of D and let its endpoints be v_i, w_j. We are no longer allowing these values to range over all of X, Y respectively. The relative contribution for this term is $O(s^{h_i}t^{h'_j}/S_{h_i}T_{h'_j})$. There are d choices for the edge in D under consideration. Therefore these choices give a relative contribution of

$$O\biggl(d\sum_{i=1}^{p}\sum_{j=1}^{q}\frac{s^{h_i}t^{h'_j}}{S_{h_i}T_{h'_j}}\biggr).$$

Combining all this together and using the bounds k, ℓ on the degrees of H in $\bar{X}$, $\bar{Y}$ respectively, we find that the required expectation is

$$\frac{\prod_{i\in\bar{X}} S_{h_i}\ \prod_{j\in\bar{Y}} T_{h'_j}}{S_{e(H)}}\left(1 + O\biggl(\frac{st}{S} + \frac{ds^{k-1}+s^k}{S_k} + \frac{dt^{\ell-1}+t^\ell}{T_\ell} + \frac{ds^k t^\ell}{S_k\,T_\ell}\biggr)\right). \qquad \square$$

For $k \in \{1, 2\}$, let $\sigma(k)$ be the expected value of the sum

$$\sum_{i=1}^{m} [d_i]_k\, [s_i - 2d_i]_{2-k}$$

when $P \in \mathcal{C}_{d,0}$ is chosen uniformly at random. Similarly define $\sigma'(k)$ to be the expected value of the sum

$$\sum_{j=1}^{n} [d'_j]_k\, [t_j - 2d'_j]_{2-k}\,.$$

These quantities will be important in the next section, so we obtain fairly precise asymptotic expressions for them below.

Lemma 3.6. *Suppose that $0 \le d \le N_2$ and that (S_2, T_2) is substantial. Then*

$$\sigma(1) = \frac{dS_3}{S_2} + O\biggl(\frac{d^2s^3t^2}{S_2T_2} + \frac{d^2s^2}{S_2} + \frac{dstS_3}{SS_2}\biggr),$$

$$\sigma'(1) = \frac{dT_3}{T_2} + O\biggl(\frac{d^2s^2t^3}{S_2T_2} + \frac{d^2t^2}{T_2} + \frac{dstT_3}{ST_2}\biggr),$$

$$\sigma(2) = \frac{[d]_2\, S_4}{{S_2}^2} + O\biggl(\frac{d^3s^4t^2}{{S_2}^2T_2} + \frac{d^3s^3}{{S_2}^2} + \frac{d^2S_4T_4}{{S_2}^2{T_2}^2} + \frac{d^2stS_4}{S{S_2}^2}\biggr),$$

$$\sigma'(2) = \frac{[d]_2\, T_4}{{T_2}^2} + O\biggl(\frac{d^3s^2t^4}{S_2{T_2}^2} + \frac{d^3t^3}{{T_2}^2} + \frac{d^2S_4T_4}{{S_2}^2{T_2}^2} + \frac{d^2stT_4}{S{T_2}^2}\biggr).$$

Proof. Let $k \in \{1, 2\}$. We will prove the formulae for $\sigma(k)$ simultaneously. The proofs for $\sigma'(k)$ are entirely analogous. Let $\mathcal{P}(k)$ be the set of all (P, v_1, v_2, w_1, w_2) satisfying the following conditions:

- $P \in \mathcal{C}_{d,0}$, $v_1, v_2 \in X$, $w_1, w_2 \in Y$,
- v_1w_1 and v_2w_2 are distinct edges of $B(P)$,
- $v_2w_2 \in D(P)$ and $v_1w_1 \notin D(P)$ if $k = 1$, while $\{v_1w_1, v_2w_2\} \subseteq D(P)$ if $k = 2$.

Let $\mathcal{Q}(k)$ be the set of all $(P, v_1, v_2, w_1, w_2) \in \mathcal{P}(k)$ such that $v_1 = v_2$. Then

$$|\mathcal{P}(k)| = |\mathcal{C}_{d,0}|\; [d]_k\, [S - 2d]_{2-k}, \qquad |\mathcal{Q}(k)| = |\mathcal{C}_{d,0}|\; \sigma(k),$$

giving

$$\sigma(k) = [d]_k\, [S - 2d]_{2-k}\, \frac{|\mathcal{Q}(k)|}{|\mathcal{P}(k)|}$$

for $k \in \{1, 2\}$.

Fix a bipartite graph D with parts X, Y and $d - k$ edges. Let $\mathcal{P}(k, D)$, respectively $\mathcal{Q}(k, D)$, be those $(P, v_1, v_2, w_1, w_2) \in \mathcal{P}(k)$, $(P, v, v, w_1, w_2) \in \mathcal{Q}(k)$ respectively, such that the nondistinguished double pairs in P correspond to D. That is, $(D, v_1, v_2, w_1, w_2) \in \mathcal{P}(1, D)$ if $D(P) = D \cup \{v_2w_2\}$ and $(P, v_1, v_2, w_1, w_2) \in \mathcal{P}(2, D)$ if $D(P) = D \cup \{v_1w_1, v_2w_2\}$, and similarly for $\mathcal{Q}(k, D)$. We will estimate $|\mathcal{Q}(k)|/|\mathcal{P}(k)|$ by finding a sufficiently accurate estimate for $|\mathcal{Q}(k, D)|/|\mathcal{P}(k, D)|$ which is independent of D.

Define $\mathcal{A}(D)$ to be the set of all (v_1, v_2, w_1, w_2) which satisfy the conditions

- $v_1, v_2 \in X$ and $w_1, w_2 \in Y$,
- if $v_1 = v_2$ then $w_1 \ne w_2$, and if $w_1 = w_2$ then $v_1 \ne v_2$,
- $v_1w_1, v_2w_2 \notin D$.

For any $(v_1, v_2, w_1, w_2) \in \mathcal{A}(D)$ let $\pi(v_1, v_2, w_1, w_2)$ be the set of ordered partitions (U_0, U_1, U_2) of $\{v_1w_1, v_2w_2\}$ into three disjoint subsets such that $v_1w_1 \notin U_2$ if $k = 1$. For $(U_0, U_1, U_2) \in \pi(v_1, v_2, w_1, w_2)$, define $n((U_0, U_1, U_2), D)$ to be the number of pairings $P' \in \cup_{j=0}^{k} \mathcal{C}_{d-j,0}$ which satisfy

- $D(P') = D \cup U_2$,
- $\sigma_{B(P')}(e) = j$ for all $e \in U_j$, for $j = 0, 1, 2$.

Then

$$\begin{aligned}
|\mathcal{P}(1,D)| &= \sum_{(v_1,v_2,w_1,w_2)\in\mathcal{A}(D)} n\big((\emptyset, \{v_1w_1\}, \{v_2w_2\}), D\big),\\
|\mathcal{P}(2,D)| &= \sum_{(v_1,v_2,w_1,w_2)\in\mathcal{A}(D)} n\big((\emptyset, \emptyset, \{v_1w_1, v_2w_2\}), D\big),\\
|\mathcal{Q}(1,D)| &= \sum_{(v,v,w_1,w_2)\in\mathcal{A}(D)} n\big((\emptyset, \{vw_1\}, \{vw_2\}), D\big),\\
|\mathcal{Q}(2,D)| &= \sum_{(v,v,w_1,w_2)\in\mathcal{A}(D)} n\big((\emptyset, \emptyset, \{vw_1, vw_2\}), D\big).
\end{aligned}$$

As a final piece of notation, let $\mathcal{C}(D) = \{P \in \mathcal{C}_{d-k,0} \mid D(P) = D\}$. Then

$$|\mathcal{C}(D)| = \sum_{(U_0,U_1,\emptyset)\in\pi(v_1,v_2,w_1,w_2)} n\big((U_0, U_1, \emptyset), D\big)$$

for any fixed $(v_1, v_2, w_1, w_2) \in \mathcal{A}(D)$. For any $(U_0, U_1, \emptyset) \in \pi(v_1, v_2, w_1, w_2)$, apply Lemma 3.2 with $L = D \cup \{v_1w_1, v_2w_2\}$, $H = 2D$ and $K = U_1$ to see that

$$|\mathcal{C}(D)| = n\big((\{v_1w_1, v_2w_2\}, \emptyset, \emptyset), D\big)\big(1 + O(st/S)\big) \tag{3.1}$$

for any fixed $(v_1, v_2, w_1, w_2) \in \mathcal{A}(D)$.

We would like to obtain an expression for the ratio $|\mathcal{Q}(k, D)|/|\mathcal{C}(D)|$. Recall the functions $f_k = f_k(D)$, $f'_k = f'_k(D)$ defined before Lemma 3.4. For all $(v, v, w_1, w_2) \in \mathcal{A}(D)$, apply Lemma 3.2 with $L = D \cup \{vw_1, vw_2\}$, $H = 2D$ and $K = kvw_1 + 2vw_2$, and use (3.1) to obtain

$$\begin{aligned}
\frac{n\big((\emptyset, \{vw_1\}, \{vw_2\}), D\big)}{|\mathcal{C}(D)|} &= \frac{[s_v - 2d_v]_3\, f'_1(w_1, w_2)}{2\,[S - 2d + 2]_3}\big(1 + O(st/S)\big),\\
\frac{n\big((\emptyset, \emptyset, \{vw_1, vw_2\}), D\big)}{|\mathcal{C}(D)|} &= \frac{[s_v - 2d_v]_4\, f'_2(w_1, w_2)}{4\,[S - 2d + 4]_4}\big(1 + O(st/S)\big),
\end{aligned}$$

for $k = 1, 2$, respectively. This gives all the terms needed for $|\mathcal{Q}(k, D)|/|\mathcal{C}(D)|$. (Since here we know that $w_1 \neq w_2$, we could write this expression without the use of the function f'_k. However it will be useful later to have the expression in this form.)

Similarly we would like an expression for $|\mathcal{P}(k, D)|/|\mathcal{C}(D)|$. Choose $(v_1, v_2, w_1, w_2) \in \mathcal{A}(D)$. Note that either of the equations $v_1 = v_2$, $w_1 = w_2$ may hold (but not both). In any case, by applying Lemma 3.2 with $H = 2D$, $K = kv_1w_1 + 2v_2w_2$ and $L = D \cup \{v_1w_1, v_2w_2\}$, and using (3.1), we obtain

$$\begin{aligned}
\frac{n\big((\emptyset, \{v_1w_1\}, \{v_2w_2\}), D\big)}{|\mathcal{C}(D)|} &= \frac{f_1(v_1, v_2)\, f'_1(w_1, w_2)}{2\,[S - 2d + 2]_3}\big(1 + O(st/S)\big),\\
\frac{n\big((\emptyset, \emptyset, \{v_1w_1, v_2w_2\}), D\big)}{|\mathcal{C}(D)|} &= \frac{f_2(v_1, v_2)\, f'_2(w_1, w_2)}{4\,[S - 2d + 4]_4}\big(1 + O(st/S)\big),
\end{aligned}$$

for $k = 1, 2$, respectively. This gives all the terms needed for $|\mathcal{P}(k, D)|/|\mathcal{C}(D)|$.

Let $D(w_1)$ be the neighbourhood of w_1 in D, and similarly for $D(w_2)$. Combining all these calculations we find that, for $k = 1, 2$,

$$\frac{|\mathcal{Q}(k, D)|}{|\mathcal{P}(k, D)|} = \frac{\sum_{\substack{w_1, w_2 \in Y \\ w_1 \neq w_2}} f'_k(w_1, w_2) \sum_{v \in X \setminus (D(w_1) \cup D(w_2))} [s_v - 2d_v]_{k+2}}{\sum_{w_1, w_2 \in Y} f'_k(w_1, w_2) \sum_{\substack{v_1 \notin D(w_1), v_2 \notin D(w_2) \\ v_1 w_1 \neq v_2 w_2}} f_k(v_1, v_2)} \big(1 + O(st/S)\big). \quad (3.2)$$

Let $S_r(D) = \sum_{v \in X} [s_v - 2d_v]_r$ for $r \geq 1$. Putting $[s_v - 2d_v]_{k+2} = O(s^{k+2})$ for each $v \in D(w_1) \cup D(w_2)$ we find

$$\sum_{v \in X \setminus (D(w_1) \cup D(w_2))} [s_v - 2d_v]_{k+2} = S_{k+2}(D) + O\big((d'_{w_1} + d'_{w_2}) s^{k+2}\big).$$

Hence the numerator of (3.2) is equal to (adding and subtracting the diagonal terms where $w_1 = w_2$):

$$\begin{aligned} &\sum_{w_1, w_2 \in Y} f'_k(w_1, w_2) \big(S_{k+2}(D) + O((d'_{w_1} + d'_{w_2}) s^{k+2})\big) + O(S_{k+2} T_{k+2}) \\ &\qquad = \sum_{w_1, w_2 \in Y} f'_k(w_1, w_2)\, S_{k+2}(D) + O(dt^k T_2 s^{k+2} + dT_k t^2 s^{k+2} + S_{k+2} T_{k+2}) \\ &\qquad = \sum_{w_1, w_2 \in Y} f'_k(w_1, w_2)\, S_{k+2}(D) + O(dT_k t^2 s^{k+2} + S_{k+2} T_{k+2}). \end{aligned}$$

Now apply Lemma 3.4(ii) to the inner summand of the denominator of (3.2) to obtain

$$\begin{aligned} &\sum_{\substack{v_1 \notin D(w_1), v_2 \notin D(w_2) \\ v_1 w_1 \neq v_2 w_2}} f_k(v_1, v_2) \\ &\qquad = S_k S_2 \left(1 + O\left(\frac{ds}{S_2} + \frac{s^{k-1}}{S_k}\right)\right) + O\big(d'_{w_1} s^k S_2 + d'_{w_2} s^2 S_k + \delta_{w_1 w_2} S_{k+2}\big), \end{aligned}$$

where $\delta_{w_1 w_2} = 1$ if $w_1 = w_2$, and $\delta_{w_1 w_2} = 0$ otherwise. Therefore the denominator of (3.2) equals

$$\begin{aligned} &\sum_{w_1, w_2 \in Y} f'_k(w_1, w_2)\, S_k S_2 \left(1 + O\left(\frac{ds}{S_2} + \frac{s^{k-1}}{S_k}\right)\right) \\ &\qquad\qquad + O\big(ds^k\, S_2 t^k T_2 + ds^2\, S_k t^2 T_k + S_{k+2} T_{k+2}\big) \\ &\quad = \sum_{w_1, w_2 \in Y} f'_k(w_1, w_2)\, S_k S_2 \left(1 + O\left(\frac{ds}{S_2} + \frac{s^{k-1}}{S_k}\right)\right) + O\big(ds^2 t^2 S_k T_k + s^2 t^2 S_k T_k\big) \\ &\quad = \sum_{w_1, w_2 \in Y} f'_k(w_1, w_2)\, S_k S_2 \left(1 + O\left(\frac{ds}{S_2} + \frac{s^{k-1}}{S_k} + \frac{ds^2 t^2}{S_2 T_2} + \frac{s^2 t^2}{S_2 T_2}\right)\right) \end{aligned}$$

using Lemma 3.4(i) for the final equality.

Now we substitute these calculations back into (3.2). In the main term, the sum over $w_1, w_2 \in Y$ cancels completely to give (using Lemma 3.4(i) and Lemma 3.1):

$$\begin{aligned}
\frac{|\mathcal{Q}(k,D)|}{|\mathcal{P}(k,D)|} &= \frac{S_{k+2}(D)}{S_kS_2}\left(1+O\Big(\frac{ds}{S_2}+\frac{s^{k-1}}{S_k}+\frac{ds^2t^2}{S_2T_2}+\frac{s^2t^2}{S_2T_2}+\frac{st}{S}\Big)\right) \\
&\qquad\qquad\qquad + O\Big(\frac{dt^2s^{k+2}}{S_kS_2T_2}+\frac{S_{k+2}T_{k+2}}{S_kS_2T_kT_2}\Big) \\
&= \frac{S_{k+2}+O(ds^{k+1})}{S_kS_2}\left(1+O\Big(\frac{ds}{S_2}+\frac{s^{k-1}}{S_k}+\frac{ds^2t^2}{S_2T_2}+\frac{s^2t^2}{S_2T_2}+\frac{st}{S}\Big)\right) \\
&\qquad\qquad\qquad + O\Big(\frac{dt^2s^{k+2}}{S_kS_2T_2}+\frac{S_{k+2}T_{k+2}}{S_kS_2T_kT_2}\Big) \\
&= \frac{S_{k+2}}{S_kS_2}\left(1+O\Big(\frac{ds}{S_2}+\frac{s^{k-1}}{S_k}+\frac{ds^2t^2}{S_2T_2}+\frac{s^2t^2}{S_2T_2}+\frac{st}{S}\Big)\right) \\
&\qquad\qquad\qquad + O\Big(\frac{dt^2s^{k+2}}{S_kS_2T_2}+\frac{ds^{k+1}}{S_kS_2}+\frac{S_{k+2}T_{k+2}}{S_lS_2T_kT_2}\Big) \\
&= \frac{S_{k+2}}{S_kS_2}+O\Big(\frac{ds^{k+2}t^2}{S_kS_2T_2}+\frac{s^{k+2}t^2}{S_kS_2T_2}+\frac{ds^{k+1}}{S_kS_2}+\frac{s^{k+1}}{S_kS_2}+\frac{S_{k+2}T_{k+2}}{S_kS_2T_kT_2}+\frac{stS_{k+2}}{S_kS_2S}\Big).
\end{aligned}$$

(We need all these error terms since $d = 0$ is possible; recall the note we made at the end of the first section.) Within the error terms given, this expression for $|\mathcal{Q}(k,D)|/|\mathcal{P}(k,D)|$ is independent of D. So the ratio $|\mathcal{Q}(k)|/|\mathcal{P}(k)|$ has the same asymptotic expression. Multiply throughout by $[d]_k\,[S-2d]_{2-k}$ to complete the proof. (In the absolute errors multiply by d^kS^{2-k}. This allows us to throw away the second and fourth error terms in the above expression, and further simplification is possible when $k=1$.) □

In the final lemma of this section we calculate two more quantities which will be used in the following section.

Lemma 3.7. *Suppose that $0 \le d \le N_2$ and that (S_2, T_2) is substantial. Choose P at random from $\mathcal{C}_{d,0}$. Then*
(i) The expected number of choices of distinct $v, x \in X$ and distinct $w, y \in Y$ such that there is a double pair from v to w and simple pairs from v to y and from x to y is

$$\frac{dS_3T_2}{SS_2}+O\Big(\frac{d^2s^3t^2}{SS_2}+\frac{dt^2S_3}{SS_2}+\frac{d^2tS_3}{SS_2}+\frac{d^2s^2T_2}{SS_2}+\frac{dstS_3T_2}{S^2S_2}+\frac{d^2s^3t^4}{SS_2T_2}\Big).$$

(ii) The expected number of choices of $v \in X$ and distinct $w, y \in Y$ such that there are simple pairs from v to w and from v to y is

$$S_2-2d-\frac{4dS_3}{S_2}+O\Big(\frac{d^2s^3t^2}{S_2T_2}+\frac{d^2s^2}{S_2}+\frac{dstS_3}{SS_2}\Big).$$

The corresponding statements hold with the roles of X and Y, s and t, S_k and T_k reversed.

Proof. First consider part (i). Let D be a fixed bipartite graph with d edges, and choose $v \in X$ and $w \in Y$ such that $vw \in D$. Take any $x \in X \setminus \{v\}$ and $y \in Y \setminus \{w\}$ such that $vy, xy \notin D$. Then by Lemma 3.3, the probability that edges xy and vy are present in a randomly chosen $P \in \mathcal{C}_{d,0}$, conditional on $D(P) = D$, is

$$\frac{(s_v - 2d_v)(s_x - 2d_x)[t_y - 2d'_y]_2}{[S - 2d]_2}\bigl(1 + O(st/S)\bigr). \tag{3.3}$$

Let $D(y)$ be the neighbourhood of y in D, and similarly $D(v)$. Now

$$\sum_{x\in X\setminus(\{v\}\cup D(y))} s_x - 2d_x = (S - 2d)\bigl(1 + O((d'_y + 1)s/S)\bigr)$$

so the sum over $x \in X \setminus (\{v\} \cup D(y))$ of (3.3) is

$$\frac{(s_v - 2d_v)[t_y - 2d'_y]_2}{S - 2d}\bigl(1 + O(d'_y s/S)\bigr)\bigl(1 + O(st/S)\bigr).$$

Also

$$\begin{aligned}\sum_{y\in Y\setminus(\{w\}\cup D(v))} [t_y - 2d'_y]_2\bigl(1 + O(d'_y s/S)\bigr) &= T_2 + O(dt) + O(dst^2/S) + O\bigl((d_v + 1)t^2\bigr)\\ &= T_2\bigl(1 + O(dt/T_2)\bigr) + O\bigl((d_v + 1)t^2\bigr)\end{aligned}$$

using Lemma 3.4(i). Therefore, for a fixed $vw \in D$, the number of x, y as above is

$$\frac{T_2(s_v - 2d_v)}{S - 2d}\bigl(1 + O(st/S + dt/T_2)\bigr) + O\bigl((d_v + 1)s_v t^2/S\bigr).$$

But for a given v, there are d_v choices for w. The required quantity is the expection for a randomly chosen $P \in \mathcal{C}_{d,0}$ of

$$\frac{T_2}{S - 2d}\sum_{v\in X} d_v(s_v - 2d_v)\bigl(1 + O(st/S + dt/T_2)\bigr) + \sum_{v\in X} O\left(\frac{d_v(d_v + 1)s_v t^2}{S}\right),$$

which is

$$\begin{aligned}&\frac{T_2}{S - 2d}\,\sigma(1)\bigl(1 + O(st/S + dt/T_2)\bigr) + O\bigl(\sigma(2)st^2/S\bigr) + O\bigl(\sigma(1)t^2/S\bigr)\\ &\quad= \frac{dS_3T_2}{SS_2} + O\left(\frac{d^2s^3t^2}{SS_2} + \frac{dt^2S_3}{SS_2} + \frac{d^2tS_3}{SS_2} + \frac{d^2s^2T_2}{SS_2} + \frac{dstS_3T_2}{S^2S_2} + \frac{d^2s^3t^4}{SS_2T_2}\right).\end{aligned}$$

The expectation in part (ii) is the expected value of

$$\sum_{v\in X} [s_v - 2d_v]_2$$

when P is chosen randomly from $\mathcal{C}_{d,0}$. We need a more accurate expression for it than that given by Lemma 3.4(i). Straightforward manipulation shows that this expectation is equal to

$$S_2 - 2d - 4\sigma(1) - 4\sigma(2).$$

Now apply Lemma 3.6. We get

$$\begin{aligned}
&S_2 - 2d - 4\sigma(1) - 4\sigma(2) \\
&\quad = S_2 - 2d - \frac{4dS_3}{S_2} + O\left(\frac{d^2s^3t^2}{S_2T_2} + \frac{d^2s^2}{S_2} + \frac{dstS_3}{SS_2}\right) \\
&\qquad + O\left(\frac{d^2S_4}{{S_2}^2} + \frac{d^3s^4t^2}{{S_2}^2T_2} + \frac{d^3s^3}{{S_2}^2} + \frac{d^2S_4T_4}{{S_2}^2{T_2}^2} + \frac{d^2stS_4}{S{S_2}^2}\right) \\
&\quad = S_2 - 2d - \frac{4dS_3}{S_2} + O\left(\frac{d^2s^3t^2}{S_2T_2} + \frac{d^2s^2}{S_2} + \frac{dstS_3}{SS_2}\right).
\end{aligned}$$

(Lemma 3.1 is used throughout to manipulate the error terms.) The proof of the final statement is entirely analogous to the above. □

4 Analysis of the switchings

We begin with a couple of technical lemmas, leading to a generalisation of [7, Lemma 7]. In the following five lemmas and corollaries, we sometimes evaluate rational functions at points where they have removable singularities. For example, in the following lemma we allow $\binom{1/B}{i}B^i$ in the case $B = 0$. In all such cases, we assume that the singularity has been removed. Thus, when $B = 0$, the function $\binom{1/B}{i}B^i$ equals $1/i!$.

Lemma 4.1. *Let j, N be integers with $N \geq 2$ and $0 \leq j \leq N$. Also let A, B and c be real numbers such that $c > 2e$, $0 \leq Ac < N - j + 1$ and $|BN| < 1$. Define*

$$\varSigma = \varSigma(A, B, N, j) = \sum_{i=0}^{N} \binom{1/B}{i} (AB)^i [i]_j.$$

Then
(i) $\varSigma = [1/B]_j (AB)^j (1 + AB)^{1/B-j} + \eta_1 (2e/c)^N [N]_j$ for some η_1 with $|\eta_1| < \frac{1}{4}$.
(ii) $\varSigma = \eta_2 [1/B]_j (AB)^j (1 + AB)^{1/B-j}$ for some η_2 with $\frac{3}{5} < \eta_2 < \frac{12}{5}$.

Proof. If $A = 0$ then (i) holds with $\eta_1 = 0$ and (ii) holds with $\eta_2 = 1$. For the remainder of the proof, assume that $A > 0$. The result for $B = 0$ holds by continuity of all our expressions with respect to B, so we may assume that $B \neq 0$. Our assumptions imply that

$$|AB| < \frac{N - j + 1}{cN} \leq \frac{N + 1}{cN} \leq \frac{3}{2c} < \frac{2}{c} < 1.$$

Therefore the infinite sum

$$S = \sum_{i=0}^{\infty} \binom{1/B}{i} (AB)^i [i]_j$$

converges and by a standard identity it is equal to $[1/B]_j (AB)^j (1 + AB)^{1/B-j}$.

Let $a_i = \binom{1/B}{i}(AB)^i[i]_j$. (Note $a_0 = 1$.) Our conditions imply that $a_i > 0$ for $0 \le i \le N$. (However, later terms can be negative.) By the continuity of all our expressions with respect to B, we can assume for simplicity that $1/B$ is not an integer. For $i \ge N$ we have

$$\frac{a_{i+1}}{a_i} = \frac{A(1 - iB)}{i - j + 1}.$$

This expression has no turning points for real i (unless it is constant), and its pole occurs for $i < N$, so its maximum value for $i \ge N$ occurs either at $i = N$ or as $i \to \infty$. This gives that $|a_{i+1}/a_i| < 2/c$ for $i \ge N$, which implies that

$$\left| \sum_{i>N} a_i \right| = a_N \left| \sum_{k=1}^{\infty} \frac{a_{N+k}}{a_N} \right| \le a_N \sum_{k=1}^{\infty} (2/c)^k \le \frac{2a_N}{c-2} \le \xi \, a_N$$

where $\xi = 1/(e-1) \sim 0.582$. Let $F = \sum_{i=0}^{N} a_i$ and $T = \sum_{i=N+1}^{\infty} a_i$. Now $F \ge a_N$ since $a_i \ge 0$ for $0 \le i \le N$, which implies that $|T/F| \le \xi$. Since $F/S = 1/(1 + T/F)$ we find that

$$0.632 < \frac{1}{1+\xi} \le \frac{F}{S} \le \frac{1}{1-\xi} < 2.393$$

and Claim (ii) follows.

To obtain (i), use the inequality $N! \ge e(N/e)^N$ to show that $a_N \le e^{-1}(2e/c)^N [N]_j$. Then it is clear that part (i) holds with $|\eta_1| \le \xi/e < \frac{1}{4}$. □

Lemma 4.2. *Let K, N be integers with $N \ge 2$ and $0 \le K \le N$. Also let A, B and c be real numbers such that $c > 2e$, $0 \le Ac < N - K + 1$ and $|BN| < 1$. Suppose that there are real numbers δ_i, for $1 \le i \le N$, and $\gamma_i \ge 0$, for $0 \le i \le K$, such that $\sum_{j=1}^{i} |\delta_j| \le \sum_{j=0}^{K} \gamma_j [i]_j < \frac{1}{5}$ for $1 \le i \le N$.*
Define $n_0, n_1, \ldots, n_N$ by $n_0 = 1$ and

$$\frac{n_i}{n_{i-1}} = \frac{A}{i}\big(1 - (i-1)B\big)\big(1 + \delta_i\big)$$

for $1 \le i \le N$, if $A \ne 0$, while $n_i = 0$ for $1 \le i \le N$ if $A = 0$. Then

$$\sum_{i=0}^{N} n_i = \exp\left(A - \tfrac{1}{2}A^2 B + \eta_4 A^3 B^2 + \eta_3 \sum_{j=0}^{K} \gamma_j (3A)^j \right) + \eta_1 (2e/c)^N$$

for some η_1, η_3, η_4 with $|\eta_1| < \frac{1}{4}$, $|\eta_3| < 4$ and $0 < \eta_4 < \frac{1}{2}$.

Proof. If $A = 0$ then the result holds with $\eta_1 = \eta_3 = \eta_4 = 0$, so we assume that $A > 0$ for the remainder of the proof. Also, the expression for n_i and the bounds we give for $\sum_{i=0}^{N} n_i$ are continuous in B, so we can also assume that $B \neq 0$.

First we prove that $\prod_{j=1}^{i}(1+\delta_j) = 1 + \theta_i \sum_{j=1}^{i} |\delta_j|$, where $-1 \le \theta_i < \frac{10}{9}$ for $1 \le i \le n$. Clearly $|\delta_j| < 1/5$ for each j, hence $\prod_{j=1}^{i}(1+\delta_j) \le \prod_{j=1}^{i}(1+|\delta_j|) \le \exp\bigl(\sum_{j=1}^{i} |\delta_j|\bigr)$. The function $(e^x - 1)/x$ is an increasing function of x for $x \ge 0$, so $\prod_{j=1}^{i}(1+\delta_j) \le 1 + \theta_i \sum_{j=1}^{i} |\delta_j|$ where $\theta_i \le 5(e^{1/5} - 1) < 10/9$. Next, note that the product is minimised, conditioned on the value of the sum, when one of the δ_j is negative and all the others are zero. Thus $\prod_{j=1}^{i}(1+\delta_j) \ge 1 - \sum_{j=1}^{i} |\delta_j|$, giving the lower bound on θ_i as claimed.

Hence we can write $n_i = a_i + b_i$ where

$$a_i = \binom{1/B}{i}(AB)^i \quad \text{and} \quad |b_i| \le \tfrac{10}{9}\binom{1/B}{i}(AB)^i \sum_{j=0}^{K} \gamma_j [i]_j.$$

By Lemma 4.1(i) with $j = 0$, we have

$$\sum_{i=0}^{N} a_i = (1 + AB)^{1/B} + \eta_1 (2e/c)^N$$

for some η_1 with $|\eta_1| < \frac{1}{4}$. Moreover, by Lemma 4.1(ii), we have

$$\sum_{i=0}^{N} |b_i| \le \tfrac{8}{3}(1 + AB)^{1/B} \sum_{j=0}^{K} \gamma_j q_j A^j,$$

where $q_j = [1/B]_j B^j (1+AB)^{-j}$. Our assumptions on A, B and N imply that $|AB| < 3/2c$, as shown in the proof of Lemma 4.1. Hence $q_{j+1}/q_j = (1-jB)/(1+AB) \le 4c/(2c-3) < 3$, so $q_j \le 3^j$ for $0 \le j < K$ (since $q_0 = 1$). Let $Q = \sum_{j=0}^{K} \gamma_j (3A)^j$. By assumption, $(3A)^j \le (Ac)^j \le (N - j + 1)^j \le [N]_j$, so $Q < \sum_{j=0}^{K} \gamma_j [N]_j < \frac{1}{5}$. Hence $\sum_{i=1}^{N} b_i = X(1 + AB)^{1/B}$ where $|X| \le 8Q/3 < 8/15$. Let $Z(x) = \log(x)/x$. Then

$$\sum_{i=0}^{N} n_i - \eta_1(2e/c)^N = (1 + AB)^{1/B}(1 + X) = (1 + AB)^{1/B}\, e^{XZ(X)} = (1 + AB)^{1/B} \exp(\eta_3 Q)$$

for some η_3 with $|\eta_3| \le 8Z(X)/3$. But $|Z(X)| \le 1.43$ when $|X| \le 8/15$, so $|\eta_3| < 4$ as claimed.

Finally, note that $0 < |AB| < 3/(2c) < 0.267$ (since we are assuming that $A, B \neq 0$). The function $R(x)$ defined by $1 + x = \exp(x - x^2/2 + R(x)x^3)$ satisfies $0 < R(x) \le R(-0.267) < 0.422$ when $0 < |x| \le 0.267$. Therefore $1 + AB = \exp\bigl(AB - \frac{1}{2}(AB)^2 + \eta_4 (AB)^3\bigr)$ where $\eta_4 = R(AB)$ satisfies $0 < \eta_4 < \frac{1}{2}$. This completes the proof. ☐

Corollary 4.3. *Suppose that there exist integers K, N and real numbers c, $\delta_1, \dots, \delta_N$, $\gamma_0, \dots, \gamma_K$ such that the requirements of Lemma 4.2 are met for all $A \in [A_1, A_2]$ and*

$B \in [B_1, B_2]$, where $0 \le A_1 \le A_2$ and $B_1 \le B_2$. Suppose $A(1), \dots, A(N) \in [A_1, A_2]$ and $B(1), \dots, B(N) \in [B_1, B_2]$. Define $n_0, n_1, \dots, n_N$ by $n_0 = 1$ and

$$\frac{n_i}{n_{i-1}} = \frac{A(i)}{i}\bigl(1 - (i-1)B(i)\bigr)\bigl(1 + \delta_i\bigr)$$

for $1 \le i \le N$, with the following interpretation: if $A(i) = 0$ then $n_j = 0$ for $i \le j \le N$. Then

$$\Sigma_1 \le \sum_{i=0}^{N} n_i \le \Sigma_2,$$

where

$$\Sigma_1 = \exp\Bigl(A_1 - \tfrac{1}{2}{A_1}^2 B_2 - 4\sum_{j=0}^{K} \gamma_j (3A_1)^j\Bigr) - \tfrac{1}{4}(2e/c)^N,$$

$$\Sigma_2 = \exp\Bigl(A_2 - \tfrac{1}{2}{A_2}^2 B_1 + \tfrac{1}{2}{A_2}^3 {B_1}^2 + 4\sum_{j=0}^{K} \gamma_j (3A_2)^j\Bigr) + \tfrac{1}{4}(2e/c)^N.$$

Proof. Define $n'_0, n'_1, \dots, n'_N$ by $n'_0 = 1$ and

$$\frac{n'_i}{n'_{i-1}} = \frac{A_1}{i}\bigl(1 - (i-1)B_2\bigr)\bigl(1 + \delta_i\bigr)$$

for $1 \le i \le N$, as before choosing $n'_i = 0$ for $i > 0$ in the case that $A_1 = 0$. Similarly define $n''_0, n''_1, \dots, n''_N$ using A_2 and B_1. From the definitions it is easy to see by induction that $0 \le n'_i \le n_i \le n''_i$ for $0 \le i \le N$, and so

$$\sum_{i=0}^{N} n'_i \le \sum_{i=0}^{N} n_i \le \sum_{i=0}^{N} n''_i.$$

Lemma 4.2 now gives the stated result. □

We will also need to apply this kind of summation argument in situations which are simpler than the above, but where one condition is weakened. We prove the necessary results by adapting the proofs of Lemmas 4.1, 4.2 and Corollary 4.3.

Lemma 4.4. *Let $N \ge 2$ be an integer, and let $\hat{c}$, A, B be real numbers such that $A \ge 0$ and $\max\{A/N,\ |AB|\} \le \hat{c} < \frac{1}{3}$. Define $n_0, n_1, \dots, n_N$ by $n_0 = 1$ and*

$$\frac{n_i}{n_{i-1}} = \frac{A}{i}\bigl(1 - (i-1)B\bigr)$$

for $1 \le i \le N$, if $A \ne 0$, and $n_i = 0$ for $1 \le i \le N$ if $A = 0$. Then

$$\sum_{i=0}^{N} n_i = \exp\bigl(A - \tfrac{1}{2}A^2 B + \psi_2 A^3 B^2\bigr) + \psi_1 (2e\hat{c})^N$$

for some ψ_1, ψ_2 with $|\psi_1| < 1$ and $0 < \psi_2 < \frac{1}{2}$.

Proof. As in Lemmas 4.1, 4.2, the cases where $A = 0$ or $B = 0$ are easily dealt with. Let $a_i = \binom{1/B}{i}(AB)^i$. For $i \geq N$ we have

$$\frac{a_{i+1}}{a_i} = \frac{A(1-iB)}{i+1} \leq \max\left\{\left|\frac{A(1-BN)}{N+1}\right|, |AB|\right\} \leq 2\hat{c}.$$

Using $N! \geq e(N/e)^N$ we have $a_N < e^{-1}(2e\hat{c})^N$. Hence

$$\left|\sum_{i>N} a_i\right| \leq a_N \sum_{k=1}^{\infty}(2\hat{c})^k \leq \frac{2\hat{c}a_N}{1-2\hat{c}} < (2e\hat{c})^N$$

since $\hat{c} < 1/3$. This gives $\sum_{i=0}^N n_i = (1+AB)^{1/B} + \psi_1(2e\hat{c})^N$ for some ψ_1 with $|\psi_1| < 1$. But as $|AB| \leq \hat{c} < \frac{1}{3}$ we have $1 + AB = \exp\bigl(AB - \frac{1}{2}(AB)^2 + \psi_2(AB)^3\bigr)$ for some ψ_2 with $0 < \psi_2 < \frac{1}{2}$, as in the proof of Lemma 4.2. $\square$

Corollary 4.5. *Let $N \geq 2$ be an integer and, for $1 \leq i \leq N$, let real numbers $A(i)$, $B(i)$ be given such that $A(i) \geq 0$ and $1 - (i-1)B(i) \geq 0$. Define $A_1 = \min_{i=1}^N A(i)$, $A_2 = \max_{i=1}^N A(i)$, $C_1 = \min_{i=1}^N A(i)B(i)$ and $C_2 = \max_{i=1}^N A(i)B(i)$. Suppose that there exists a real number $\hat{c}$ with $0 < \hat{c} < \frac{1}{3}$ such that* $\max\{A/N, |C|\} \leq \hat{c}$ *for all $A \in [A_1, A_2]$, $C \in [C_1, C_2]$. Define $n_0, \ldots, n_N$ by $n_0 = 1$ and*

$$\frac{n_i}{n_{i-1}} = \frac{A(i)}{i}\bigl(1-(i-1)B(i)\bigr)$$

for $1 \leq i \leq N$, with the following interpretation: if $A(i) = 0$ or $1 - (i-1)B(i) = 0$, then $n_j = 0$ for $i \leq j \leq N$. Then

$$\Sigma_1 \leq \sum_{i=0}^N n_i \leq \Sigma_2$$

where

$$\Sigma_1 = \exp\bigl(A_1 - \tfrac{1}{2}A_1C_2\bigr) - (2e\hat{c})^N,$$
$$\Sigma_2 = \exp\bigl(A_2 - \tfrac{1}{2}A_2C_1 + \tfrac{1}{2}A_2{C_1}^2\bigr) + (2e\hat{c})^N.$$

Proof. First we prove the upper bound. If $A_2 = 0$ then it is easy to verify that the conclusion holds. Otherwise define $n_i'' = \binom{A_2/C_1}{i}{C_1}^i$ for $0 \leq i \leq N$. (Recall that this is defined even when $C_1 = 0$.) By induction on i, $n_i \leq n_i''$ for $0 \leq i \leq N$, so $\sum_{i=0}^N n_i \leq \sum_{i=0}^N n_i''$. Applying Lemma 4.4 to the last sum (with $A = A_2$, $B = C_1/A_2$) gives the upper bound Σ_2, as required.

For the lower bound, define $n_i' = \binom{A_1/C_2}{i}{C_2}^i$ for $0 \leq i \leq N$. If $A_1 = 0$ then $\Sigma_1 = 1 - (2e\hat{c})^N$, which is a lower bound since $n_0 = 1$. We may now assume that $A_1 > 0$. By continuity of our expressions for n_i' and Σ_1 with respect to C_2, we may assume that $C_2 \neq 0$.

Suppose first that there exists $j \le N$ such that $n'_j < 0$. We cannot invoke Lemma 4.4 immediately since it applies only to non-negative series. Instead, define $\ell_0 = 1$ and for $i = 1, 2, \ldots, N$, let

$$\ell_i = \begin{cases} n'_i & \text{if } n'_i > 0 \text{ and } \ell_{i-1} > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Then $\ell_i \le n_i$ for $0 \le i \le N$ by induction, so $\sum_{i=0}^{N} n_i \ge \sum_{i=0}^{N} \ell_i$. Let $f(x) = (1+x)^y$ for real x, y. Using Taylor's theorem with remainder,

$$f(x) = \sum_{i=0}^{k-1} \binom{y}{i} x^i + \frac{f^{(k)}(\xi) x^k}{k!} = \sum_{i=0}^{k-1} \binom{y}{i} x^i + \binom{y}{k} x^k (1+\xi)^{y-k},$$

where $\xi = \xi(x,y) \in (0,x)$ if $x > 0$ and $\xi \in (x,0)$ if $x < 0$. Provided that $x > -1$, it follows that the tail of the Taylor expansion starting from the kth term has the same sign as the kth term. Recall that $C_2 \ne 0$, and that $|C_2| < 1/3$, which implies that $C_2 > -1$. Substituting $x = C_2$ and $y = A_1/C_2$ gives

$$\sum_{i=0}^{N} n_i \ge \sum_{i=0}^{N} \ell_i \ge \sum_{i=0}^{\infty} n'_i.$$

However, as in the proof of Lemma 4.4,

$$\sum_{i=0}^{\infty} n'_i = (1+C_2)^{A_1/C_2} \ge \exp\bigl(A_1 - \tfrac{1}{2} A_1 C_2\bigr).$$

This expression is bounded below by Σ_1, as required.

Finally suppose that $n'_j \ge 0$ for $0 \le j \le N$. Then

$$\sum_{i=0}^{N} n_i \ge \sum_{i=0}^{N} \ell_i = \sum_{i=0}^{N} n'_i.$$

Applying Lemma 4.4 to the right hand side (with $A = A_1$, $B = C_2/A_1$) gives the lower bound Σ_1. □

We can now use switchings to estimate the relative sizes of some of the classes $\mathcal{C}_{d,h}$.

Lemma 4.6. *Suppose* $0 \le d \le N_2$ *and* $1 < h \le N_3$, *with* $|\mathcal{C}_{d,h}| \ne 0$. *If* (S_2, T_2) *is substantial then*

$$\frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{d,h-1}|} = \frac{S_3 T_3 + O\bigl(s^2 t^2 (st + d + h) S\bigr)}{6 h S^3}.$$

Proof. Note, if $|\mathcal{C}_{d,h}| \neq 0$ then $|\mathcal{C}_{d,h-1}| \neq 0$ so the left hand ratio is well defined. Choose an arbitrary $P \in \mathcal{C}_{d,h}$. Define $N = N(P)$ to be the number of t-switchings which can be applied to P. We can choose a triple pair and its labelling in $6h$ ways, and choose three distinct labelled simple pairs (p_4, p'_4), (p_5, p'_5) and (p_6, p'_6) in $[S - 2d - 3h]_3$ ways (in the notation of Figure 1). Unwanted coincidences like $v(p_1) = v(p_4)$ or $v(p_4) = v(p_5)$ account for $O(h(s+t)S^2)$ choices. The forbidden cases where, for example, P already has a pair involving $v(p_1)$ and $v(p'_4)$ account for $O\big(hstS^2\big)$ choices. Overall, we find that

$$N = 6hS^3\big(1 + O((st + d + h)/S)\big).$$

Now choose an arbitrary $P' \in \mathcal{C}_{d,h-1}$, and let $N' = N'(P)$ be the number of inverse t-switchings which can be applied to it. We can choose two distinct 3-stars of simple pairs (one star centred in X, the other in Y) in

$$S_3T_3 - O\big((d+h)(s^2T_3 + t^2S_3)\big)$$

ways. Of these choices, we must eliminate those not permitted. An unwanted coincidence of a pair from each star occurs in at most $O(s^2t^2S)$ choices. An unwanted additional pair, such as from $v(p_1)$ to $v(p'_1)$ or $v(p_4)$ to $v(p'_4)$ occurs in at most $O(s^3t^3S)$ choices. Hence

$$N' = S_3T_3 + O\big(s^2t^2(st + d + h)S\big).$$

The lemma follows on considering the ratio N'/N. □

Corollary 4.7. *Suppose $0 \leq d \leq N_2$ with $|\mathcal{C}_{d,0}| \neq 0$. Further suppose that (S_2, T_2) is substantial. Then*

$$\sum_{h=0}^{N_3} \frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{d,0}|} = \exp\left(\frac{S_3T_3}{6S^3} + O\big(s^2t^2(st+d)/S^2\big)\right).$$

Proof. We will apply Corollary 4.5. Let h' be the first value of $h \leq N_3$ for which $|\mathcal{C}_{d,h}| = 0$, or $h' = N_3 + 1$ if there is no such value. Define α_h, $1 \leq h < h'$, by

$$\frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{d,h-1}|} = \frac{S_3T_3 - \alpha_h\big(s^2t^2(st + d + (h-1)S)\big)}{6hS^3}. \tag{4.1}$$

Lemma 4.6 says that α_h is bounded independently of h, d and S.

For $1 \leq h < h'$, define

$$A(h) = \frac{S_3T_3 - \alpha_h(s^2t^2(st+d)S)}{6S^3}, \quad C(h) = \frac{\alpha_h s^2t^2}{6S^2}.$$

If $\alpha_h \leq 0$ then $A(h) > 0$ by its definition. (We can't have $A(h) = 0$ because of the assumption that $h < h'$.) If $\alpha_h > 0$ then $C(h) > 0$, which implies that $A(h) > 0$ since the right side of (4.1) is $A(h) - (h-1)C(h)$. Therefore $A(h) > 0$ whenever $h < h'$. Define $B(h) = C(h)/A(h)$ for $1 \le h < h'$. Also define $A(h) = B(h) = 0$ for $h' \le h \le N_3$.

Define A_1, A_2, C_1, C_2 by taking the minimum and maximum of the $A(h)$ and $C(h)$ over $1 \le h \le N_3$, as in Corollary 4.5. Let $A \in [A_1, A_2]$ and $C \in [C_1, C_2]$, and set $\hat{c} = \frac{1}{41}$. Since $A = S_3T_3/(6S^3) + o(1)$ and $C = o(1)$, we have that $\max\{A/N_3, |C|\} < \hat{c}$ for S sufficiently large, by the definition of N_3.

Therefore Corollary 4.5 applies and says that

$$\sum_{h=0}^{N_3} \frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{d,0}|} = \exp\biggl(\frac{S_3T_3}{6S^3} + O\bigl(s^2t^2(st+d)/S^2\bigr)\biggr) + O\bigl((2e/41)^{N_3}\bigr).$$

Finally, $(2e/41)^{N_3} \le (2e/41)^{\log S} \le S^{-2}$. Since the sum we are estimating is at least equal to one, this additive error term is covered by the error terms inside the exponential. This completes the proof. □

Now we turn our attention to the distribution of the number of double pairs, in pairings with no pairs of multiplicity greater than 2.

Lemma 4.8. *Suppose that (S_2, T_2) is substantial and that $1 \le d \le N_2$ with $|\mathcal{C}_{d,0}| \ne 0$. Then*

$$\frac{|\mathcal{C}_{d,0}|}{|\mathcal{C}_{d-1,0}|} = \frac{A(d)}{d}\bigl(1 - (d-1)B\bigr)(1+\delta_d)$$

where

$$\begin{aligned}
A(d) &= \frac{S_2T_2}{2S^2}\biggl(1 + \frac{S_2}{S^2} + \frac{T_2}{S^2} + \frac{1}{S} + \frac{2S_3T_2}{S_2S^2} + \frac{2S_2T_3}{S^2T_2} - \frac{S_3T_3}{SS_2T_2} - \frac{2S_2T_2}{S^3}\biggr) + O\biggl(\frac{s^3t^3}{S^2}\biggr),\\
B &= \frac{2}{S_2} + \frac{2}{T_2} + \frac{4T_3}{{T_2}^2} + \frac{4S_3}{{S_2}^2} - \frac{4}{S},\\
\delta_d &= O\biggl(\frac{(d-1)^2s^2}{{S_2}^2} + \frac{(d-1)^2t^2}{{T_2}^2} + \frac{dst(d+st)}{S_2T_2}\biggr).
\end{aligned}$$

Proof. Note that, if $|\mathcal{C}_{d,0}| \ne 0$ then $|\mathcal{C}_{d-1,0}| \ne 0$, so the left hand ratio is well defined. We will use the notation of Figure 1. In addition, e_i is the pair (p_i, p_i'), for $i = 1, \ldots, 4$. Lemma 3.1 is used throughout to simplify error terms.

Let N be the number of available d-switchings for a random $P \in \mathcal{C}_{d,0}$; precisely, the expected number of tuples (e_1, e_2, e_3, e_4) satisfying all the requirements for a d-switching.

First, denote by X_1 the class of choices of (e_1, e_2, e_3, e_4) such that e_1 and e_2 are distinct parallel double pairs, e_3 and e_4 are simple pairs, and the six cells $\{v(p_1), v(p_3), v(p_4), v(p_1'), v(p_3'), v(p_4')\}$ are distinct. Having chosen $P \in \mathcal{C}_{d,0}$, we can choose (e_1, e_2) in $2d$ ways, then two distinct simple pairs e_3 and e_4 in $[S-2d]_2$ ways. From this we must subtract the choices where $v(p_1) = v(p_3)$. By Lemma 3.6, on average these number

$$2(S-2d-1)\sigma(1) = \frac{2dSS_3}{S_2} + O\biggl(\frac{d^2s^3t^2S}{S_2T_2} + \frac{d^2s^2S}{S_2} + \frac{dstS_3}{S_2}\biggr).$$

(The factor of 2 accounts for distinguishing between the two edges of the double pair in 2 ways.) The choices where $v(p_1) = v(p_4)$ have the same average count, whereas the choices for each of the possibilities $v(p'_1) = v(p'_3)$ and $v(p'_1) = v(p'_4)$ have an average count

$$\frac{2dST_3}{T_2} + O\biggl(\frac{d^2s^2t^3S}{S_2T_2} + \frac{d^2t^2S}{T_2} + \frac{dstT_3}{T_2}\biggr),$$

by symmetry. In addition, it might be that $v(p_3) = v(p_4)$. This has an average count of

$$2dS_2 - 4d^2 - \frac{8d^2S_3}{S_2} + O\biggl(\frac{d^3s^3t^2}{S_2T_2} + \frac{d^3s^2}{S_2} + \frac{d^2stS_3}{SS_2}\biggr)$$

by Lemma 3.7(ii), and in the same way the possibility $v(p'_3) = v(p'_4)$ has an average count of

$$2dT_2 - 4d^2 - \frac{8d^2T_3}{T_2} + O\biggl(\frac{d^3s^2t^3}{S_2T_2} + \frac{d^3t^2}{T_2} + \frac{d^2stT_3}{ST_2}\biggr).$$

We have enumerated six possible coincidences. If any two of them occur simultaneously, we have a maximum count less than $O(d(s+t)^2)$ by just counting the cases. Combining these estimates, we find that the average size of X_1 is

$$\begin{aligned}
&2d[S-2d]_2 - \frac{4dSS_3}{S_2} - \frac{4dST_3}{T_2} - 2dS_2 - 2dT_2 \\
&\quad + O\biggl(d^2 + \frac{d^2S_3}{S_2} + \frac{d^2T_3}{T_2} + \frac{d^2s^2S}{S_2} + \frac{d^2s^3t^2S}{S_2T_2} + \frac{dstS_3}{S_2} + \frac{d^2t^2S}{T_2} + \frac{d^2s^2t^3S}{S_2T_2} \\
&\qquad\qquad + \frac{dstT_3}{T_2} + d(s+t)^2\biggr) \\
&= 2dS^2 - \frac{4dSS_3}{S_2} - \frac{4dST_3}{T_2} - 2dS_2 - 2dT_2 - 8d^2S - 2dS \\
&\quad + O\biggl(\frac{d^2s^2S}{S_2} + \frac{d^2s^3t^2S}{S_2T_2} + \frac{dstS_3}{S_2} + \frac{d^2t^2S}{T_2} + \frac{d^2s^2t^3S}{S_2T_2} + \frac{dstT_3}{T_2} + d(s+t)^2 + d^3\biggr) \\
&= 2dS^2 - \frac{4dSS_3}{S_2} - \frac{4dST_3}{T_2} - 2dS_2 - 2dT_2 - 8d^2S - 2dS \\
&\quad + O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr).
\end{aligned}$$

Some of the choices in X_1 are not valid for d-switchings because there are already pairs (simple or double) from $v(p_1)$ to $v(p'_3)$ or $v(p'_4)$, or pairs from $v(p'_1)$ to $v(p_3)$ or $v(p_4)$. Let X_2 be the subset of X_1 which has this difficulty. By Lemma 3.7(i), there are on average

$$\begin{aligned}
&2\bigl(S - 2d - O(s+t)\bigr) \\
&\qquad \times \biggl(\frac{dS_3T_2}{SS_2} + O\biggl(\frac{d^2s^3t^2}{SS_2} + \frac{d^2s^2T_2}{SS_2} + \frac{dt^2S_3}{SS_2} + \frac{d^2tS_3}{SS_2} + \frac{d^2s^3t^4}{SS_2T_2} + \frac{dstS_3T_2}{S^2S_2}\biggr)\biggr) \\
&\quad = \frac{2dS_3T_2}{S_2} + O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr)
\end{aligned}$$

choices that give a simple pair from $v(p_1)$ to $v(p'_3)$, and the same number that give a simple pair from $v(p_1)$ to $v(p'_4)$. (Again the factor of 2 comes from labelling p_1 and p_2 in two different ways.) Similarly, the other two undesired simple pairs each give counts

$$\frac{2dS_2T_3}{T_2} + O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr).$$

Two of these four possibilities occur together for $O(ds^2t^2)$ choices, on average. (This follows from direct counting and using Lemma 3.2 where necessary.) A double pair from $v(p_1)$ to $v(p'_3)$ occurs for

$$O\bigl(tS\sigma(2)\bigr) = O\biggl(\frac{d^2tS_4S}{{S_2}^2} + \frac{d^3s^3tS}{{S_2}^2} + \frac{d^3s^4t^3S}{{S_2}^2T_2}\biggr) = O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr)$$

choices, by Lemma 3.6. Similarly if there is a double pair from $v(p_1)$ to $v(p'_4)$, while double pairs from $v(p_3)$ or $v(p_4)$ to $v(p'_1)$ occur for

$$O\bigl(sS\sigma'(2)\bigr) = O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr)$$

choices. Combining these estimates we find that the average size of X_2 is

$$\frac{4dS_3T_2}{S_2} + \frac{4dS_2T_3}{T_2} + O\biggl(\frac{d^2st(d+st)S^2}{S_2T_2} + ds^2t^2\biggr).$$

Putting all this together gives

$$\begin{aligned} N = 2dS^2 \biggl(1 &- \frac{2S_3}{SS_2} - \frac{2T_3}{ST_2} - \frac{S_2}{S^2} - \frac{T_2}{S^2} - \frac{4d}{S} - \frac{1}{S} - \frac{2S_3T_2}{S_2S^2} - \frac{2S_2T_3}{T_2S^2} \\ &+ O\biggl(\frac{dst(d+st)}{S_2T_2} + \frac{s^2t^2}{S^2}\biggr)\biggr). \end{aligned}$$

Now we must consider inverse d-switchings. With reference to Figure 1, define $e_1 = (p_1, p'_3)$, $e_2 = (p_2, p'_4)$, $e_3 = (p_3, p'_1)$, $e_4 = (p_4, p'_2)$. Let N' be the number of available inverse d-switchings for a random $P \in \mathcal{C}_{d-1,0}$; precisely, the expected number of tuples (e_1, e_2, e_3, e_4) satisfying all the requirements for an inverse d-switching.

We begin with the set Y_1 of choices (e_1, e_2, e_3, e_4) of simple pairs with the six cells $\{v(p_1), v(p_3), v(p_4), v(p'_1), v(p'_3), v(p'_4)\}$ distinct. The pairs (e_1, e_2) and (e_3, e_4) can be chosen independently in

$$\begin{aligned} &\biggl(S_2 - 2(d-1) - \frac{4(d-1)S_3}{S_2}\biggr)\biggl(T_2 - 2(d-1) - \frac{4(d-1)T_3}{T_2}\biggr) \\ &\qquad + O\biggl(\frac{(d-1)^2s^2T_2}{S_2} + \frac{(d-1)^2t^2S_2}{T_2} + \frac{d^2s^3t^2}{S_2} + \frac{d^2s^2t^3}{T_2} + \frac{dstS_3T_2}{SS_2} + + \frac{dstS_2T_3}{ST_2}\biggr) \\ &= S_2T_2 - 2(d-1)\biggl(S_2 + T_2 + \frac{2S_2T_3}{T_2} + \frac{2S_3T_2}{S_2}\biggr) \\ &\qquad + O\biggl(\frac{(d-1)^2s^2T_2}{S_2} + \frac{(d-1)^2t^2S_2}{T_2} + st(d+st)^2\biggr) \end{aligned}$$

ways, by Lemma 3.7(ii). (Keeping $d-1$ instead of d in the first two terms of the error will be significant, which is why we do not simplify these here.) From these we subtract the choices where $\{e_1, e_2, e_3, e_4\}$ are not distinct: on average these number

$$\frac{4S_2T_2}{S} + O\biggl(\frac{stS_2T_2}{S^2} + \frac{dsT_2}{S} + \frac{s^2T_2}{S} + \frac{dtS_2}{S} + \frac{t^2S_2}{S} + \frac{ds^2t^2}{S}\biggr) = \frac{4S_2T_2}{S} + O\bigl((d+st)st\bigr)$$

by Lemma 3.5. We also subtract the choices where $v(p_1) = v(p_3)$ but $\{e_1, e_2, e_3, e_4\}$ are distinct, and the three similar cases. First consider the possibility that $v(p_1) = v(p_3)$. If $(d-1)s^2 + s^3 = o(S_3)$ then Lemma 3.5 applies and says that there are

$$\frac{S_3T_2}{S} + O\biggl(\frac{stS_3T_2}{S^2} + \frac{ds^2T_2}{S} + \frac{s^3T_2}{S} + \frac{dtS_3}{S} + \frac{t^2S_3}{S} + \frac{ds^3t^2}{S}\biggr) = \frac{S_3T_2}{S} + O\bigl((d+st)s^2t^2\bigr)$$

such choices. However if the condition fails then either $S_3 = O(ds^2)$ and there are at most $O(ds^2t)$ choices, or $S_3 = O(s^3)$ and there are at most $O(s^3t)$ such choices. But these counts are both covered by the given error term. The same estimate holds for the number of choices with $v(p_1) = v(p_4)$, while for each of the situations that $v(p'_1) = v(p'_3)$ and $v(p'_1) = v(p'_4)$ the estimate is

$$\frac{S_2T_3}{S} + O\bigl((d+st)s^2t^2\bigr).$$

These exceptions are disjoint, so we have that the average size of Y_1 is

$$S_2T_2 - \frac{4S_2T_2}{S} - \frac{2S_3T_2}{S} - \frac{2S_2T_3}{S} - 2(d-1)\left(S_2 + T_2 + \frac{2S_2T_3}{T_2} + \frac{2S_3T_2}{S_2}\right)$$
$$+ O\biggl(\frac{(d-1)^2s^2T_2}{S_2} + \frac{(d-1)^2t^2S_2}{T_2} + st(d+st)^2\biggr).$$

Within the choices Y_1, a subset Y_2 do not give legal inverse d-switchings because there is a pair from $v(p_1)$ to $v(p'_1)$, from $v(p_3)$ to $v(p'_3)$, or from $v(p_4)$ to $v(p'_4)$. If $(d-1)s^2+s^3 = o(S_3)$ and $(d-1)t^2 + t^3 = o(T_3)$ then Lemma 3.5 applies and says that that number of choices with a simple pair from $v(p_1)$ to $v(p'_1)$ is

$$\frac{S_3T_3}{S} + O\biggl(\frac{stS_3T_3}{S^2} + \frac{ds^2T_3}{S} + \frac{s^3T_3}{S} + \frac{dt^2S_3}{S} + \frac{t^3S_3}{S} + \frac{ds^3t^3}{S}\biggr)$$
$$= \frac{S_3T_3}{S} + O\bigl((d+st)s^2t^2\bigr)$$

choices. On the other hand, if $S_3 = O(ds^2)$ or $T_3 = O(dt^2)$ then the number of such choices is $O(ds^2t^2)$, while if $S_3 = O(s^3)$ then the number is at most $O(s^3t^2)$ and similarly $O(s^2t^3)$ if $T_3 = O(t^3)$. These counts are covered by the stated error term. The average number of choices in Y_1 where there is a simple pair from $v(p_3)$ to $v(p'_3)$, say, is

$$\frac{{S_2}^2{T_2}^2}{S^3} + O\biggl(\frac{st{S_2}^2{T_2}^2}{S^4} + \frac{dsS_2{T_2}^2}{S^3} + \frac{s^2S_2{T_2}^2}{S^3} + \frac{dt{S_2}^2T_2}{S^3} + \frac{t^2{S_2}^2T_2}{S^3} + \frac{ds^2t^2S_2T_2}{S^3}\biggr)$$
$$= \frac{{S_2}^2{T_2}^2}{S^3} + O\bigl((d+st)s^2t^2\bigr),$$

by Lemma 3.5, and there are the same number of choices with a simple pair from $v(p_4)$ to $v(p_4')$, on average. Two of these possibilities may occur together for $O(s^2t^2S_2T_2/S^2)$ choices, on average (this can be proved using Lemma 3.2). A double pair in one of the three forbidden positions occurs for $O(ds^2t^2)$ choices. Thus, the average size of Y_2 is

$$\frac{S_3T_3}{S} + \frac{2{S_2}^2{T_2}^2}{S^3} + O\big((d+st)s^2t^2\big).$$

Combining these estimates, we find that

$$\begin{aligned} N' = S_2T_2 &- \frac{4S_2T_2}{S} - \frac{2S_3T_2}{S} - \frac{2S_2T_3}{S} - \frac{S_3T_3}{S} - \frac{2{S_2}^2{T_2}^2}{S^3} \\ &- 2(d-1)\left(S_2 + T_2 + \frac{2S_2T_3}{T_2} + \frac{2T_2S_3}{S_2}\right) \\ &+ O\left(\frac{(d-1)^2s^2T_2}{S_2} + \frac{(d-1)^2t^2S_2}{T_2} + st(d+st)^2\right). \end{aligned}$$

Writing

$$N = 2dS^2\big(1 - \Delta + O(\varepsilon)\big), \quad N' = S_2T_2(1-\Delta') + O(\varepsilon'),$$

we find that

$$\frac{N'}{N} = \frac{S_2T_2}{2dS^2}\big(1 - \Delta' + \Delta + O(\Delta\Delta' + \Delta^2 + \varepsilon)\big) + O\left(\frac{\varepsilon'}{dS^2}\right)$$

where

$$\begin{aligned} \Delta &= \frac{2S_3}{SS_2} + \frac{2T_3}{ST_2} + \frac{S_2}{S^2} + \frac{T_2}{S^2} + \frac{4d}{S} + \frac{1}{S} + \frac{2S_3T_2}{S^2S_2} + \frac{2S_2T_3}{S^2T_2}, \\ \Delta' &= \frac{4}{S} + \frac{2S_3}{SS_2} + \frac{2T_3}{ST_2} + \frac{S_3T_3}{SS_2T_2} + \frac{2S_2T_2}{S^3} + 2(d-1)\left(\frac{1}{T_2} + \frac{1}{S_2} + \frac{2T_3}{{T_2}^2} + \frac{2S_3}{{S_2}^2}\right), \\ \varepsilon &= \frac{dst(d+st)}{S_2T_2} + \frac{s^2t^2}{S^2}, \\ \varepsilon' &= \frac{(d-1)^2s^2T_2}{S_2} + \frac{(d-1)^2t^2S_2}{T_2} + st(d+st)^2. \end{aligned}$$

Note that $\Delta, \Delta', \varepsilon = o(1)$ and $\Delta\Delta' + \Delta^2 = O(\varepsilon)$, though this is somewhat tedious to verify. Next,

$$\begin{aligned} \Delta - \Delta' = (d-1)&\left(\frac{4}{S} - \frac{2}{T_2} - \frac{2}{S_2} - \frac{4T_3}{{T_2}^2} - \frac{4S_3}{{S_2}^2}\right) \\ &+ \frac{S_2}{S^2} + \frac{T_2}{S^2} + \frac{1}{S} + \frac{2S_3T_2}{S_2S^2} + \frac{2S_2T_3}{S^2T_2} - \frac{S_3T_3}{SS_2T_2} - \frac{2S_2T_2}{S^3}. \end{aligned}$$

From here it is not difficult to check that the statement of the lemma holds. □

Corollary 4.9. *If (S_2, T_2) is substantial then*

$$\sum_{d=0}^{N_2}\sum_{h=0}^{N_3}\frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{0,0}|}$$
$$= \exp\biggl(\frac{S_2T_2}{2S^2}+\frac{S_2T_2}{2S^3}-\frac{S_3T_3}{3S^3}+\frac{S_2T_2(S_2+T_2)}{4S^4}+\frac{{S_2}^2T_3+S_3{T_2}^2}{2S^4}-\frac{{S_2}^2{T_2}^2}{2S^5}+O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\biggr).$$

Proof. We need to apply Corollary 4.3 to the result of Lemma 4.8, and take into account the terms coming from the triples (as given by Corollary 4.7).

Let d' be the first value of $d \le N_2$ for which $|\mathcal{C}_{d,0}| = 0$, or $d' = N_2 + 1$ if no such value of d exists. Define $m_0, m_1, \ldots, m_{N_2}$ by

$$m_d = \frac{|\mathcal{C}_{d,0}|}{|\mathcal{C}_{0,0}|}\sum_{h=0}^{N_3}\frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{d,0}|}$$

for $0 \le d < d'$, and $m_d = 0$ for $d' \le d \le N_2$. Then clearly

$$\sum_{d=0}^{N_2}\sum_{h=0}^{N_3}\frac{|\mathcal{C}_{d,h}|}{|\mathcal{C}_{0,0}|} = \sum_{d=0}^{N_2} m_d.$$

Corollary 4.7 tells us that for $d < d'$ we have

$$m_d = \frac{|\mathcal{C}_{d,0}|}{|\mathcal{C}_{0,0}|}\exp\biggl(\frac{S_3T_3}{6S^3}+O(s^3t^3/S^2)+\xi_d s^2t^2/S^2\biggr)$$

where $\xi_0 = 0$ and in general $\xi_d = O(d)$. (Note that this inequality is also true for $d' \le d \le N_2$, since both sides equal zero.) If α is a constant such that $|\xi_d| \le \alpha d$ for $0 \le d \le d'$, then

$$\exp\biggl(\frac{S_3T_3}{6S^3}+O(s^3t^3/S^2)\biggr)\sum_{d=0}^{N_2} n_d(-1) \le \sum_{d=0}^{N_2} m_d \le \exp\biggl(\frac{S_3T_3}{6S^3}+O(s^3t^3/S^2)\biggr)\sum_{d=0}^{N_2} n_d(1) \tag{4.2}$$

where

$$n_d(x) = \frac{|\mathcal{C}_{d,0}|}{|\mathcal{C}_{0,0}|}\exp\bigl(x\alpha d s^2t^2/S^2\bigr).$$

Next we note that, for $x \in \{-1, 1\}$, $n_0(x) = 1$, and for $1 \le d \le d'$,

$$\frac{n_d(x)}{n_{d-1}(x)} = A(d)\bigl(1-(d-1)B\bigr)\bigl(1+\delta_d\bigr)$$

with $A(d)$, B, and δ_d satisfying the expressions given in the statement of Lemma 4.8. This follows since the factor $\exp(x\alpha s^2t^2/S^2)$ is covered by the error term on $A(d)$. For $d' \le d \le N_2$ define $A(d) = 0$.

Now let $A_1 = A_1(x) = \min_d A(d)$, $A_2 = A_2(x) = \max_d A(d)$, where the maximum and minimum are taken over $1 \le d \le N_2$. Also let $B_1 = B_2 = B$, and $K = 3$, and define $c = S^{1/4}$ if $N_2 = 8$ and $c = 41$ otherwise. The conditions of Corollary 4.3 now hold as we will show. Let $A \in [A_1, A_2]$ be arbitrary.

Clearly $c > 2e$. If $N_2 = 8$ then $S_2T_2 < S^{7/4}$ and $Ac = 1/2(1+o(1)) < N_2 - 2$, and otherwise $Ac = 41S_2T_2/(2S^2)(1+o(1)) < 21S_2T_2/S^2 - 2 \le N_2 - 2$. It is easy to check that $BN_2 = o(1)$ so for S large enough we have $|BN_2| < 1$. If $d = O(S_2T_2/S^2)$ then

$$\sum_{d=1}^{N_2} |\delta_d| = O\left(\frac{s^2 S_2 {T_2}^3}{S^6} + \frac{t^2 {S_2}^3 T_2}{S^6} + \frac{st {S_2}^2 {T_2}^2}{S^6} + \frac{s^2 t^2 S_2 T_2}{S^4}\right) = O\left(\frac{s^3 t^3}{S^2}\right) = o(1),$$

while if $d \le \lceil \log S \rceil$ then

$$\sum_{d=1}^{N_2} |\delta_d| = O\left(\frac{s^2 \log^3 S}{{S_2}^2} + \frac{t^2 \log^3 S}{{T_2}^2} + \frac{st \log^3 S}{S_2 T_2} + \frac{s^2 t^2 \log^2 S}{S_2 T_2}\right) = o(1).$$

Finally, for $1 \le k \le N_2$, we have

$$\begin{aligned}\sum_{d=1}^{k} |\delta_d| &= O\left(\sum_{d=1}^{k} (d-1)^2 \left(\frac{s^2}{{S_2}^2} + \frac{t^2}{{T_2}^2}\right)\right) + O\left(\sum_{d=1}^{k} \frac{d^2 st}{S_2 T_2}\right) + O\left(\sum_{d=1}^{k} \frac{d s^2 t^2}{S_2 T_2}\right) \\ &= O\left(k(k-1)(2k-1)\left(\frac{s^2}{{S_2}^2} + \frac{t^2}{{T_2}^2}\right) + \frac{k(k+1)(2k+1)st}{S_2 T_2} + \frac{k(k+1)s^2 t^2}{S_2 T_2}\right) \\ &= \sum_{j=0}^{K} \gamma_j [k]_j,\end{aligned}$$

where

$$\gamma_0 = 0,\ \gamma_1 = O\left(\frac{s^2 t^2}{S_2 T_2}\right),\ \gamma_2 = O\left(\frac{s^2}{{S_2}^2} + \frac{t^2}{{T_2}^2} + \frac{s^2 t^2}{S_2 T_2}\right),\ \gamma_3 = O\left(\frac{s^2}{{S_2}^2} + \frac{t^2}{{T_2}^2} + \frac{st}{S_2 T_2}\right).$$

Since ${N_2}^3(s^2/{S_2}^2 + t^2/{T_2}^2 + st/S_2T_2) = o(1)$, which is easily checked, it follows that $\sum_{j=0}^{K} \gamma_j [k]_j < 1/5$ for $1 \le k \le N_2$, when S is large enough.

Therefore the conditions of Corollary 4.3 hold, and we conclude that each of the bounds given by that Corollary for $\sum_{d=0}^{N_2} n_d(x)$ has the form

$$\exp\left(A - \frac{A^2 B}{2} + O\left(A^3 B^2 + \sum_{j=0}^{3} \gamma_j (3A)^j\right)\right) + O\left((2e/c)^{N_2}\right),$$

where A is either A_1 or A_2. A somewhat tedious check shows that

$$O(A^3 B^2) + \sum_{j=0}^{3} \gamma_j (3A)^j = O(s^3 t^3 / S^2).$$

Next consider the error term $O\big((2e/c)^{N_2}\big)$. If $N_2 = 8$ then $(2e/c)^{N_2} = (2eS^{-1/4})^8 = O(S^{-2})$, while in the other cases we have $(2e/c)^{N_2} = (2e/41)^{N_2} \le (2e/41)^{\log S} = O(S^{-2})$, by our choice of c. Since $n_0 = 1$, this additive error term is covered by a relative error of the same form. Therefore, each of the bounds on $\sum_{d=0}^{N_2} n_d(x)$ has the form

$$\exp\biggl(A - \frac{A^2 B}{2} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\biggr) = \exp\biggl(\frac{S_2T_2}{2S^2} + \frac{S_2T_2}{2S^3} - \frac{S_3T_3}{2S^3} + \frac{S_2T_2(S_2+T_2)}{4S^4} + \frac{S_3{T_2}^2 + {S_2}^2T_3}{2S^4} - \frac{{S_2}^2{T_2}^2}{2S^5} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\biggr).$$

Modulo the given error terms, the final expression does not depend on x, nor on whether we are taking a lower bound or upper bound in Corollary 4.3. To complete the proof, just apply (4.2). ☐

We now have the proof of our main theorem.

Proof of Theorem 1.3. If $1 \le st = o(S^{2/3})$ and (S_2, T_2) is not substantial then the result holds, by Lemma 2.2. On the other hand, if (S_2, T_2) is substantial, the result follows from Corollary 4.9 and Lemma 2.3. ☐

5 Alternative formulation

For some applications, Theorem 1.3 is not in a very convenient form. We now give another formulation. For $k = 2, 3$, define

$$\mu_k = \frac{mn}{S(mn-S)} \sum_{i=1}^{m} (s_i - S/m)^k$$
$$\nu_k = \frac{mn}{S(mn-S)} \sum_{j=1}^{n} (t_j - S/n)^k.$$

To motivate the definitions, recall that S/m is the mean value of s_i and S/n is the mean value of t_j, so these are scaled central moments.

Corollary 5.1. *Under the conditions of Theorem 1.3,*

$$N(\boldsymbol{s},\boldsymbol{t}) = \frac{\prod_{i=1}^{m}\binom{n}{s_i}\prod_{j=1}^{n}\binom{m}{t_j}}{\binom{mn}{S}} \exp\biggl(-(1-\mu_2)(1-\nu_2)\Bigl(\frac{1}{2} + \frac{1+\mu_2\nu_2}{4S}\Bigr) + \frac{(1-\mu_2)(1-\mu_2+2\mu_2\nu_2)}{4n} + \frac{(1-\nu_2)(1-\nu_2+2\mu_2\nu_2)}{4m} + \frac{(1-3{\mu_2}^2+2\mu_3)(1-3{\nu_2}^2+2\nu_3)}{12S} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\biggr).$$

Proof. By Stirling's formula or otherwise,

$$\binom{N}{x} = \frac{N^x}{x!} \exp\Bigl(-\frac{[x]_2}{2N} - \frac{[x]_3}{6N^2} - \frac{[x]_2}{4N^2} + O(x^4/N^3)\Bigr)$$

as $N \to \infty$, provided the error term is bounded. This gives us the approximations

$$\prod_{i=1}^{m} \binom{n}{s_i} = \frac{n^S}{\prod_i s_i!} \exp\Bigl(-\frac{S_2}{2n} - \frac{S_2}{4n^2} - \frac{S_3}{6n^2} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\Bigr)$$

$$\prod_{j=1}^{n} \binom{m}{t_j} = \frac{m^S}{\prod_j t_j!} \exp\Bigl(-\frac{T_2}{2m} - \frac{T_2}{4m^2} - \frac{T_3}{6m^2} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\Bigr)$$

$$\binom{mn}{S} = \frac{(mn)^S}{S!} \exp\Bigl(-\frac{S^2}{2mn} + \frac{S}{2mn} - \frac{S^3}{6m^2n^2} + O\Bigl(\frac{s^3t^3}{S^2}\Bigr)\Bigr).$$

Substitute these expressions into Theorem 1.3 and replace S_2, S_3, T_2, T_3 by their equivalents in terms of $\mu_2, \mu_3, \nu_2, \nu_3$. The desired result is obtained. □

Most of the terms inside the exponential of Corollary 5.1 are tiny unless μ_2 and/or ν_2 are quite large (that is, the graph is very far from semiregular). For example, we have the following simplification.

Corollary 5.2. *If the conditions of Theorem 1.3 hold and also* $(1+\mu_2)(1+\nu_2) = O(S^{1/3})$, *then*

$$N(\boldsymbol{s}, \boldsymbol{t}) = \frac{\prod_{i=1}^{m} \binom{n}{s_i} \prod_{j=1}^{n} \binom{m}{t_j}}{\binom{mn}{S}} \exp\bigl(-\tfrac{1}{2}(1-\mu_2)(1-\nu_2) + O(st/S^{2/3})\bigr).$$

Proof. It is only necessary to check that the additional terms in Corollary 5.1 have the required size. It helps to realise that $\mu_2 \le s$, $\mu_3 \le s\mu_2$, $\nu_2 \le t$ and $\nu_2 \le t\nu_2$. □

Form a random bipartite graph (with m vertices in one part of the bipartition and n in the other) by independently placing an edge in each of the mn available positions with probability S/mn. Standard calculations show that the expected values of μ_2 and ν_2 are exactly 1, while the expected values of μ_3 and ν_3 equal $1-2S/mn$, which is $1-o(1)$ under our assumptions. In a future paper, we will show that the argument of the exponential in Corollary 5.2 is vanishing in this case with high probability. This will allow us to apply the result easily to the degree distributions of random bipartite graphs.

References

[1] A. Békéssy, P. Békéssy and J. Komlós, Asymptotic enumeration of regular matrices, *Studia Sci. Math. Hungar.*, **7** (1972) 343–353.

[2] E. A. Bender, The asymptotic number of nonnegative integer matrices with given row and column sums, *Discrete Math.*, **10** (1974) 345–353.

[3] B. Bollobás and B. D. McKay, The number of matchings in random regular graphs and bipartite graphs, *J. Combinatorial Theory (Series B)*, **41** (1986) 80–91.

[4] E. R. Canfield and B. D. McKay, Asymptotic enumeration of 0-1 matrices with constant row and column sums, *Electron. J. Combin.*, to appear (2005).

[5] C. J. Everett, Jr., and P. R. Stein, The asymptotic number of integer stochastic matrices, *Discrete Math.*, **1** (1971) 33–72.

[6] B. D. McKay, Asymptotics for 0-1 matrices with prescribed line sums, *Enumeration and Design, Academic Press, Canada*, (1984) 225–238.

[7] B. D. McKay and X. Wang, Asymptotic enumeration of 0-1 matrices with equal row sums and equal column sums, *Linear Alg. Appl.*, **373** (2003) 273–288.

[8] B. D. McKay and N. C. Wormald, Uniform generation of random regular graphs of moderate degree, *J. Algorithms* **11** (1990) 52–67.

[9] B. D. McKay and N. C. Wormald, Asymptotic enumeration by degree sequence of graphs with degrees $o(n^{1/2})$, *Combinatorica*, **11** (1991) 369–382.

[10] M. P. Mineev and A. I. Pavlov, On the number of (0,1)-matrices with prescribed sums of rows and columns, *Doc. Akad. Nauk SSSR 230*, **2** (1976) 1276–1282.

[11] P. E. O'Neil, Asymptotics and random matrices with row-sum and column-sum restrictions, *Bull. Amer. Math. Soc.*, **75** (1969) 1276–1282.

[12] R. C. Read, Some enumeration problems in graph theory, *Doctoral Thesis*, *London University*, (1958).

[13] X. Wang, Asymptotic enumeration of bipartite graphs, tournaments, digraphs and Eulerian digraphs with multiple edges, Ph.D. thesis, Australian National University, (1995).

[14] N. C. Wormald, Some problems in the enumeration of labelled graphs, *Doctoral Thesis*, *Dept. of Maths., University of Newcastle*, (1978).