Asymptotics for 0−1 Matrices With Prescribed Line Sums

Brendan D. McKay

Australian National University
Canberra, Australia

ABSTRACT

Let $M(n, k)$ be the number of $n \times n$ 0−1 matrices with each line sum equal to k. Let $0 < \epsilon < \frac{1}{6}$. Then

$$M(n, k) = \frac{(nk)!}{(k!)^{2n}} \exp\left(-\frac{(k-1)^2}{2} + 0\left(\frac{k^3}{n}\right)\right)$$

uniformly for $1 \le k \le \epsilon n$. This is generalised to rectangular 0−1 matrices with arbitrary (possibly non-equal) line sums. A limited set of compulsory zeroes can also be specified.

1. Introduction.

Let $g = g(n) = (g_1, g_2, \ldots, g_n)$ and $g' = g'(n') = (g_{1'}, g_{2'}, \ldots, g_{n'})$ be sequences of non-negative integers, and define $M(g, g')$ to be the set of all 0−1 matrices of order $n \times n'$ whose i-th row sum is g_i $(1 \le i \le n)$ and whose i-th column sum is $g_{i'}$ $(1 \le i \le n')$. For obvious reasons, we will always assume that $\sum_{i=1}^{n} g_i = \sum_{i=1}^{n'} g_{i'}$. We will be concerned with the asymptotic properties of $M(g, g')$ as $n, n' \to \infty$. In particular, let $N(g, g') = |M(g, g')|$. The value of $N(g, g')$ has been the object of much study over the past 20 years, so we start by summarizing the previous results.

Define $I = \{1, 2, \ldots, n\}$, $I' = \{1', 2', \ldots, n'\}$, $I^* = I \cup I'$, $g_{\max} = \max_{i \in I^*} g_i$, and $\alpha = 2(\sum_{i=1}^{n} \binom{g_i}{2}) (\sum_{i=1}^{n'} \binom{g_{i'}}{2})\, e(G)^{-2}$, where $e(G) = \sum_{i \in I} g_i$. Define $P(g, g')$ by

ISBN 0-12-379120-0

$$N(g, g') = \frac{e(G)!}{\prod_{i \in I^*} g_i!} P(g, g').$$

The first result of interest to us was that of Read [11], who proved that $P(g, g') = e^{-\alpha} + o(1)$ if $g_i = 3$ for all $i \in I^*$. The same result with fixed constant line sums was established by Everett and Stein [5] and extended to arbitrary but bounded line sums by Békéssy, Békéssy and Komlós [1], Bender [2] and Wormald [13]. The first attempt to allow $g_{\max}$ to increase with $n + n'$ was by O'Neil [10], who proved that $P(g, g') = e^{-\alpha}(1 + 0(n^{-1+\delta}))$, provided the row and column sums are (independently) equal and $g_{\max} \le (\log n)^{1/4-\epsilon}$. This was improved by Mineev and Pavlov [9], who obtained the following:

(i) If $n = n'$ and all line sums are k, then $P(g, g') = e^{-\alpha} + 0(n^{-1+\gamma^2/2})$, provided $1 < k \le \gamma \log^{1/2} n$ and $0 < \gamma < 1$ (γ fixed).

(ii) If all row sums are k and all column sums are $l > 1$ then $P(g, g') = e^{-\alpha} + 0(\log^3 n/n^{1-\gamma})$, provided $1 < k \le (l-1)^{-1} \gamma \log n$ and $0 < \gamma < 1$ (γ fixed).

(iii) $P(g, g') = e^{-\alpha} + 0(n^{\gamma^4/4-1/2}\log^2 n)$ if $g_{\max} \le \gamma \log^{1/4} n$ and $0 < \gamma < (2/3)^{1/4}$ (γ fixed).

Most recently, Bollobás and McKay [4] have shown that $P(g, g') = e^{-\alpha}(1 + 0(n^{-3/4}))$ if all line sums are k and $k = 0(\log^{1/3} n)$.

Actually, some of the results quoted above have more generality. [1] and [2], for example, optionally allow integer entries greater than one. Also, [10] (to some extent), [2], [13] and [4] prescribe a suitably restricted set of matrix entries which must be zero. This extension enables one to investigate such things as the expected number of submatrices of specified form ([10], [3], [4], [12]), a matter which has also been studied by other means in [6].

Leaving aside differences of terminology, all the papers cited above, except [6], use essentially the same model; only the method and accuracy of the analysis varies. This paper is no exception. Our contribution is a new method of analysis which enables us to considerably extend and strengthen all previous results.

2. The Model.

Consider a collection of disjoint sets v_i ($i \in I^*$), where v_i has cardinality g_i. These will henceforth be called **cells.** A **pairing** is a set of pairs (called the **edges** of the pairing) such that

(i) each edge has the form (x, x'), where $x \in \bigcup_{i \in I} v_i$ and

$x' \in \bigcup_{i \in I'} v_i$, and

(ii) each element of $\bigcup_{i \in I^*} v_i$ is in exactly one edge.

Given a pairing **P**, we can obtain a bipartite multigraph $G(\mathbf{P})$. The two parts of the multigraph are $\{v_i \mid i \in I\}$ and $\{v_i \mid i \in I'\}$. The number of graph edges joining v_i to $v_{i'}$ is the number of edges (x, x') of **P** such that $x \in v_i$ and $x' \in v_{i'}$. Clearly, vertex v_i has degree g_i in $G(\mathbf{P})$. Equivalently, **P** yields an $n \times n'$ integer matrix whose (i, i')-th entry equals the number of edges in $G(\mathbf{P})$ from v_i to $v_{i'}$. For the remainder of the paper we will use the graph terminology rather than the matrix terminology.

Given g and g', the number of possible pairings is clearly $e(G)!$. Furthermore, each element of $\mathbf{M}(g, g')$ corresponds to exactly $\Pi_{i \in I^*} g_i!$ pairings. $P(g, g')$ can thus be interpreted as the probability that a randomly chosen pairing produces a multigraph with no multiple edges. Previous estimates of $P(g, g')$ were made using either inclusion-exclusion or the method of moments. In the next section we present a method which is more complex yet substantially more accurate.

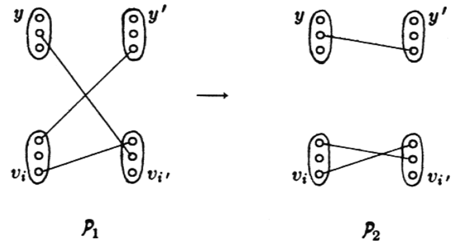
3. Basic Analysis.

For notational convenience, we will make no distinction between graphs and their edge sets. Also, define $V = \{v_i \mid i \in I\}$ and $V' = \{v_i \mid i \in I'\}$. We will always assume that $g_{\max} \geq 1$.

Let L be a simple bipartite graph with parts V and V'. Let $l_{\max}$ be the maximum degree of L, and let $e(L)$ be the number of edges. Furthermore, let H be a bipartite multigraph on the same parts as L, with the restriction that if xx' is an edge of non-zero multiplicity of H, then xx' is an edge of L. Let h_i be the degree of vertex v_i in H and let $e(H) = \sum_{i \in I} h_i$ be the total number of edges. If $x \in V$ and $x' \in V'$ then $\mu_H(xx')$ denotes the multiplicity of the edge xx' in H. Also, $H + xx'$ is the multigraph obtained from H by adding an extra edge from x to x'.

Let $\mathbf{C}(L, H)$ be the set of all pairings **P** such that, for $x \in V$ and $x' \in V'$, if xx' is an edge of L then $\mu_{G(\mathbf{P})}(xx') = \mu_H(xx')$ and if xx' is not an edge of L then $\mu_{G(\mathbf{P})}(xx') \leq 1$. More informally, $G(\mathbf{P})$ is simple outside L and coincides with H inside L.

Let $\mathbf{P}_1 \in \mathbf{C}(L, H)$ and $\mathbf{P}_2 \in \mathbf{C}(L, H + v_i v_{i'})$, where $v_i v_{i'} \in L$. Then $\mathbf{P}_1$ and $\mathbf{P}_2$ are said to be **closely related** if $\mathbf{P}_2$ can be obtained from $\mathbf{P}_1$ by an operation of the following sort.



In the diagram, only the relevant parts of $\mathbf{P}_1$ and $\mathbf{P}_2$ are drawn; the other parts are unchanged by the operation. In detail, the requirements on the operation are

(i) $v_i, y \in V$; $v_{i'}, y' \in V'$; $v_i \neq y$, $v_{i'} \neq y'$,

(ii) $v_i v_{i'} \in L$, and

(iii) $v_i y', y v_{i'}, yy' \notin L$.

Define $N_1 = |\mathbf{C}(L, H)|$ and $N_2 = |\mathbf{C}(L, H + v_i v_{i'})|$. Furthermore, let M be the number of closely related pairs $(\mathbf{P}_1, \mathbf{P}_2)$ such that $\mathbf{P}_1 \in \mathbf{C}(L, H)$ and $\mathbf{P}_2 \in \mathbf{C}(L, H + v_i v_{i'})$. Our aim for the remainder of this section will be to bound the ratio of N_1 to N_2. We will do this via four separate bounds on M.

3.1 LEMMA. $M \leq (g_i - h_i)(g_{i'} - h_{i'})N_1$.

PROOF: Given $\mathbf{P}_1 \in \mathbf{C}(L, H)$, y can be chosen in at most $g_i - h_i$ ways and y' in at most $g_{i'} - h_{i'}$ ways.

3.2 LEMMA. Let $\Delta = 2g_{\max}(g_{\max} + l_{\max})$. Then

$$M \geq (e(G) - e(H) - \Delta)(\mu_H(v_i v_{i'}) + 1)\, N_2.$$

PROOF: Choose $\mathbf{P}_2 \in \mathbf{C}(L, H + v_i v_{i'})$. We wish to bound the number of pairings $\mathbf{P}_1 \in \mathbf{C}(L, H)$ which are closely related to $\mathbf{P}_2$. The edge from v_i to $v_{i'}$ can be chosen in $\mu_H(v_i v_{i'}) + 1$ ways. To choose the edge from y to y' we consider all the possibilities. To begin with, there are $e(G) - e(H) - 1$ edges of $\mathbf{P}_2$ which lie outside $H + v_i v_{i'}$. Not all of these are eligible, however. The possible things which can go wrong are listed below, together with the maximum numbers of edges which can be excluded by each.

$$v_i = y : g_i - h_i - 1$$

$$v_{i'} = y' : g_{i'} - h_{i'} - 1$$

$v_i y' \in G(\mathbf{P}_2) : (g_i - \mu_H(v_i\, v_{i'}) - 1)(g_{\max} - 1)$

$v_{i'} y \in G(\mathbf{P}_2) : (g_{i'} - \mu_H(v_i\, v_{i'}) - 1)(g_{\max} - 1)$

$v_i y' \in L \backslash G(\mathbf{P}_2) : (l_{\max} - 1)(g_{\max} - 1)$

$v_{i'} y \in L \backslash G(\mathbf{P}_2) : (l_{\max} - 1)(g_{\max} - 1)$

The Lemma can now be proved by subtracting the sum of these numbers from $e(G) - e(H) - 1$.

Lemmas 3.1 and 3.2 can be combined to obtain our first bound on the ratio of N_1 to N_2.

3.3 THEOREM. If $e(G) - e(H) - \Delta \geq 1$ and $N_1 \neq 0$, then

$$\frac{N_2}{N_1} \leq \frac{(g_i - h_i)(g_{i'} - h_{i'})}{(\mu_H(v_i\, v_{i'}) + 1)(e(G) - e(H) - \Delta)}.$$

Let J be a bipartite multigraph which satisfies the same requirements as H, and define $e(J)$ and $\{j_i\}$ consistently with $e(H)$ and $\{h_i\}$. Let $H + J$ be the bipartite multigraph such that $\mu_{H+J}(xx') = \mu_H(xx') + \mu_J(xx')$ for all $x \in V$ and $x' \in V'$. By applying Theorem 3.3 the right number of times we obtain the following extension. For integers a, b define $a^{[b]} = a(a-1)\cdots(a-b+1)$.

3.4 THEOREM. If $e(G) - e(H) - \Delta \geq e(J)$ and $\mathbf{C}(L, H) \neq \emptyset$, then

$$\frac{|\mathbf{C}(L, H+J)|}{|\mathbf{C}(L, H)|} \leq \frac{\Pi_{i \in I^*}(g_i - h_i)^{[j_i]}}{(e(G) - e(H) - \Delta)^{[e(J)]}\Pi_{xx'}\,(\mu_{H+J}(xx'))^{[\mu_J(xx')]}},$$

where the product in the denominator is over all $x \in V$ and $x' \in V'$.

Our next task is to obtain bounds complementary to Theorems 3.3 and 3.4. The first prerequisite follows easily from the proof of Lemma 3.2.

3.5 LEMMA. $M \leq (e(G) - e(H) - 1)(\mu_H(v_i\, v_{i'}) + 1)N_2$.

The other prerequisite is not so easy to come by.

3.6 LEMMA. Let $\Delta' = 2 + 2g_{\max}(g_{\max} + l_{\max} + 2)$, and suppose that $e(G) - e(H) - \Delta' > 0$. Then

$$M \geq (g_i - h_i)(g_{i'} - h_{i'})\left(1 - \frac{g_{\max}(g_{\max} + l_{\max})}{e(G) - e(H) - \Delta'}\right)N_1.$$

PROOF: Unfortunately, there is no useful lower bound on the number of pairings $\mathbf{P}_2 \in \mathbf{C}(L, H + v_i\, v_{i'})$ which are closely related to a particular $\mathbf{P}_1 \in \mathbf{C}(L, H)$. Instead, we will choose a random

$\mathbf{P}_1 \in \mathbf{C}(L, H)$ and bound the *expected* number of closely related pairings in $\mathbf{C}(L, H + v_i v_{i'})$.

The upper bound given in Lemma 3.1 is high because of the possibilities that $yy' \in L$ or $\mu_{G(\mathbf{P}_1)}(yy') \neq 0$. We will consider the second possibility first.

Choose a random $\mathbf{P}_1 \in \mathbf{C}(L, H)$ and then a random $y \in V$ such that $yv_{i'} \in G(\mathbf{P}_1)\backslash L$. The latter choice can be made in $g_{i'} - h_{i'}$ ways. Given y, we can choose $y' \in V'$, such that $yy' \in G(\mathbf{P}_1)$, in at most $g_{\max} - 1$ ways. If now $v_i y' \in L$ or $y' = v_{i'}$ we have nothing to worry about. Otherwise, the probability that $v_i y' \in G(\mathbf{P}_1)$ is at most

$$\frac{|\mathbf{C}(L \cup \{yv_{i'}, yy', v_i y'\}, H + yv_{i'} + yy' + v_i y')|}{|\mathbf{C}(L \cup \{yv_{i'}, yy', v_i y'\}, H + yv_{i'} + yy')|}$$

$$\leq \frac{(g_i - h_i)(g_{\max} - 1)}{e(G) - e(H) - 2 - \Delta_1}, \text{ where } \Delta_1 = 2g_{\max}(g_{\max} + l_{\max} + 2),$$

by Theorem 3.3. Thus the number of cases excluded by the possibility that $\mu_{G(\mathbf{P}_1)}(yy') \neq 0$ is at most

$$\frac{(g_i - h_i)(g_{i'} - h_{i'})(g_{\max} - 1)^2 N_1}{e(G) - e(H) - 2 - \Delta_1}.$$

The second potential problem is the possibility that $yy' \in L\backslash G(\mathbf{P}_1)$. To investigate this, choose a random $\mathbf{P}_1 \in \mathbf{C}(L, H)$ and then a random $y \in V$ such that $yv_{i'} \in G(\mathbf{P}_1)\backslash L$, as before. Then choose $y' \in V'$ such that $yy' \in L\backslash G(\mathbf{P}_1)$. This can be done in at most $l_{\max}$ ways. If $y' = v_{i'}$ or $v_i y' \in L$ there is no problem, so suppose otherwise. Then the probability that $v_i y' \in G(\mathbf{P}_1)$ is at most

$$\frac{|\mathbf{C}(L \cup \{yv_{i'}, v_i y'\}, H + yv_{i'} + v_i y')|}{|\mathbf{C}(L \cup \{yv_{i'}, v_i y'\}, H + yv_{i'})|}$$

$$\leq \frac{(g_i - h_i)g_{\max}}{e(G) - e(H) - 1 - \Delta_2}, \text{ where } \Delta_2 - 2g_{\max}(g_{\max} + l_{\max} + 1),$$

by Theorem 3.3. Thus the number of cases excluded by the possibility $yy' \in L\backslash G(\mathbf{P}_1)$ is at most

$$\frac{(g_i - h_i)(g_{i'} - h_{i'})g_{\max} l_{\max} N_1}{e(G) - e(H) - 1 - \Delta_2}.$$

The Lemma is now obtained by adding these two bounds.

We can now obtain bounds complementary to Theorems 3.3 and 3.4.

3.7 THEOREM. If $e(G) - e(H) - \Delta' \geq 1$ and $N_1 \neq 0$, then

$$\frac{N_2}{N_1} \geq \frac{(g_i - h_i)\,(g_{i'} - h_{i'})}{(\mu_H(v_i\, v_{i'}) + 1)(e(G) - e(H) - 1)} \left(1 - \frac{g_{\max}(g_{\max} - l_{\max})}{e(G) - e(H) - \Delta'}\right).$$

3.8 THEOREM. If $e(G) - e(H) - \Delta' > e(J)$ and $\mathbf{C}(L, H) \neq \varnothing$, then

$$\frac{|\mathbf{C}(L, H+J)|}{|\mathbf{C}(L, H)|} \leq \frac{\prod_{i \in I^*} (g_i - h_i)^{[j_i]}}{(e(G) - e(H) - 1)^{[e(J)]} \prod_{xx'} (\mu_{H+J}(xx'))^{[\mu_J(xx')]}} \times \left(1 - \frac{g_{\max}(g_{\max} + l_{\max})}{e(G) - e(H) - e(J) - \Delta'}\right)^{e(J)},$$

where the product in the denominator is over all xx' such that $x \in V$ and $x' \in V'$.

For later convenience, we summarize Theorems 3.4 and 3.8 for the special case where H has no edges.

3.9 THEOREM. If $e(G) - \Delta' > e(J)$ and $\mathbf{C}(L, \varnothing) \neq \varnothing$, then

$$\frac{|\mathbf{C}(L, J)|}{|\mathbf{C}(L, \varnothing)|} = \frac{\prod_{i \in I^*} g_i^{[j_i]}}{(e(G) - 1)^{[e(J)]} \prod_{xx'} \mu_J(xx')!} D(g, g', L, J),$$

where

$$\left(1 - \frac{g_{\max}(g_{\max} + l_{\max})}{e(G) - e(J) - \Delta'}\right)^{e(J)} \leq D(g, g', L, J) \leq \frac{(e(G) - 1)^{[e(J)]}}{(e(G) - \Delta)^{[e(J)]}}.$$

4. SYNTHESIS

The results of the previous section can be used to find an estimate of the probability $P(g, g')$. In fact we will do this with a little more generality. Let X be simple bipartite graph with parts V and V'. Define $P = P(g, g', X)$ to be the probability that a random pairing $\mathbf{P}$ produces a simple graph $G(\mathbf{P})$ with no edges in common with X. In the matrix formulation, X specifies a set of matrix entries which must be zero.

Choosing a pairing $\mathbf{P}$ and form the multigraph $G(\mathbf{P})$. A **naughty** edge of $G(\mathbf{P})$ is one that is either parallel to another edge (i.e. is a part of a multiple edge) or coincides with an edge of X (or both). The naughty edges of $G(\mathbf{P})$ together form a multigraph called the **naughty**

graph of $G(\mathbf{P})$ (and of $\mathbf{P}$). Our problem is to estimate the probability P that a random $\mathbf{P}$ has a naughty graph with no edges. By definition, this is equal to the number of pairings with an empty naughty graph divided by the total number. Thus we have

$$\frac{1}{P} = \sum_K \frac{\nu(K)}{\nu(\emptyset)}$$

if $P \neq 0$, where $\nu(K)$ is the number of pairings with naughty graph K, and the sum is over all possible K. The ratio $\nu(K)/\nu(\emptyset)$ can be written in terms of the ratio bounded in Theorem 3.9. To do this, separate each possible naughty graph K into edge-disjoint multigraphs $A(K)$ and $B(K)$. The edges of $A(K)$ are those edges of K which coincide with edges X, and the edges of $B(K)$ are those which don't. Then clearly

$$\frac{\nu(K)}{\nu(\emptyset)} = \frac{|\mathbf{C}(X \cup B(K),K)|/|\mathbf{C}(X \cup B(K),\emptyset)|}{\sum_{S \subseteq B(K)} |\mathbf{C}(X \cup B(K),S)|/|\mathbf{C}(X \cup B(K), \emptyset)|},$$

where the sum is over all simple subgraphs S of $B(K)$.

The value of $1/P$ can now be estimated by comparing it to a more tractible expression. For $i \in I$, $i' \in I'$ and $k \geq 0$ define

$$\alpha_k(ii') = \frac{g_i^{[k]} g_{i'}^{[k]}}{k!e(G)^k},$$

and

$$\Psi = \Psi(g, g', X) = \prod_{ii' \in X} (1 + \alpha_1(ii') + \alpha_2(ii')) \prod_{ii' \notin X} (1 + \alpha_2(ii')).$$

When the products are expanded and multiplied, they yield a summation some terms of which can be identified with a possible naughty graph K. Precisely, K can be identified with the term

$$\psi(K) = \prod_{ii' \in K} \alpha_{\mu_K(ii')}(ii'),$$

which is present if and only if K has no edges of multiplicity greater than two. There may also be terms which would correspond to graphs K with one or more vertices of too large a degree.

We now present a series of lemmas which will enable us to compare the values of Ψ and $1/P$.

4.1 LEMMA. Let K be a possible naughty graph. Then, if $\nu(\emptyset) \neq \emptyset$,

$$\sum_{L \supset K} \frac{\nu(L)}{\nu(\varnothing)} \le \frac{1}{Pe(G)^{[e(K)]}} \prod_{ii'} \frac{g_i^{[\mu_k(ii')]} g_{i'}^{[\mu_k(ii')]}}{\mu_K(ii')!},$$

where the sum is over all possible naughty graphs L which contain K, and the product is over all $ii' \in V \times V'$.

PROOF. Since the total number of pairings is $e(G)!$, we have

$$\sum_{L \supset K} \frac{\nu(L)}{\nu(\varnothing)} = \frac{1}{P} \sum_{L \supset K} \frac{\nu(L)}{e(G)!}.$$

The product in the Lemma bounds the number of ways of choosing $e(K)$ pairing edges which correspond to K. The remaining pairing edges can be chosen in $(e(G) - e(K))!$ ways.

4.2 LEMMA. Let K be a partite multigraph with parts V and V'. Then

$$\sum_{L \supset K} \psi(L) \le \psi(L)\Psi,$$

where the sum is over all bipartite multigraphs which contain K and have the same parts as K.

PROOF: This is immediate from the definition of Ψ.

For a possible naughty graph K, define

$$r(K) = \frac{\prod_{ii' \in K} g_i^{[\mu_K(ii')]} g_{i'}^{[\mu_K(ii')]}}{\prod_{i \in I^*} g_i^{[k_i]}}.$$

Note that $r(K) \ge 1$. Also, define $\hat{\Delta} = 3 + 2g_{\max}(2g_{\max} + x_{\max} + 2)$.

4.3 LEMMA. Let K be a possible naughty graph with no edges of multiplicity greater than two. Assume that $\hat{\Delta} \le \epsilon_1 e(G)$ and $e(K) \le \epsilon_2 e(G)$, where ϵ_1 and ϵ_2 are fixed positive constants with $\frac{3}{2}\epsilon_1 + \epsilon_2 < 1$. Then, if $\nu(\varnothing) \ne \varnothing$,

$$1 + \left| \frac{\nu(K)/\nu(\varnothing)}{\psi(K)} - 1 \right| \le r(K) \exp\left(0\left(\frac{\hat{\Delta} e(K)}{e(G)}\right)\right) \frac{e(G)^{e(K)}}{e(G)^{[e(K)]}}.$$

PROOF: By Theorem 3.9,

$$1 + \left| \frac{\nu(K)/\nu(\varnothing)}{\psi(K)} - 1 \right| \le r(K) \frac{e(G)^{e(K)}}{(e(G) - \hat{\Delta})^{[e(K)]}}$$

$$\times \left(1 - \frac{\hat{\Delta}/2}{e(G) - e(K) - \hat{\Delta}}\right)^{-e(K)} \sum_{S \supseteq B(K)} \frac{|\mathbf{C}(X + B(K), S)|}{|\mathbf{C}(X + B(K), \varnothing)|}$$

where the sum is over all simple subgraphs S of $B(K)$.

The assumed bounds on $\hat{\Delta}$ and $e(K)$ ensure that

$$\frac{e(G)^{e(K)}}{(e(G) - \hat{\Delta})^{[e(K)]}} = \exp\left(0\left(\frac{e(K)\hat{\Delta}}{e(G)}\right)\right) \frac{e(G)^{e(K)}}{e(G)^{[e(K)]}}$$

and

$$\left(1 - \frac{\hat{\Delta}/2}{e(G) - e(K) - \hat{\Delta}}\right)^{-e(K)} = \exp\left(0\left(\frac{e(K)\hat{\Delta}}{e(G)}\right)\right)$$

To bound the sum, let m be the number of edges of $B(K)$, not counting multiplicities. Obviously $m \leq e(K)/2$. The number of simple subgraphs S of $B(K)$ with exactly r edges is at most $\binom{m}{r}$. Thus, using Theorem 3.9, the sum is bounded by

$$\sum_{r=0}^{m} \binom{m}{r} \frac{g_{\max}^{2r}}{(e(G) - \hat{\Delta})^{[r]}} \leq \left(1 + \frac{e g_{\max}^{2}}{e(G) - \hat{\Delta}}\right)^{m}$$

$$= \exp\left(0\left(\frac{\hat{\Delta} e(K)}{e(G)}\right)\right)$$

4.4 LEMMA. Suppose that $\hat{\Delta} \leq \epsilon_1 e(G)$, where $\epsilon_1 < \frac{2}{3}$. Then

$$\frac{1}{P} = \Psi \exp\left(0\left(\frac{\hat{\Delta}^2}{e(G)}\right)\right).$$

PROOF: We proceed by breaking the problem into a number of pieces. Firstly, choose a constant ϵ_2 such that $0 < \epsilon_2 < \frac{2}{3} - \epsilon_1$. We can dispose of those K with $e(K) > \epsilon_2 e(G)$ by arguing as follows.

The function

$$f(x) = \frac{1}{\Psi} \prod_{ii' \in X} (1 + \alpha_1(ii')x + \alpha_2(ii')x^2) \prod_{ii' \notin X} (1 + \alpha_2(ii')x^2)$$

can be interpreted as the probability generating function of a non-negative random variable which is the sum of nn' simpler independent random variables. The mean of the random variable is bounded by

$$\sum_{ii'} \frac{g_i^{[2]} g_{i'}^{[2]}}{e(G)^2} + \sum_{ii' \epsilon X} \frac{g_i g_{i'}}{e(G)} \leq \frac{\hat{\Delta}}{2}.$$

Therefore, the contribution to Ψ of those K with $e(K) > \epsilon_2 e(G)$ is at most $\hat{\Delta}\Psi/(2\epsilon_2 e(G)) = \Psi 0(\hat{\Delta}/e(G))$. A parallel argument (working directly with pairings) shows that the contribution to $1/P$ of those K with $e(K) > \epsilon_2 e(G)$ is at most $0(\hat{\Delta}/e(G))/P$.

At this point, it is convenient to note that the displayed bound above guarantees the existence of at least one actually occurring naughty graph K for which $e(K) \leq \hat{\Delta}/2$. Because $\epsilon_1 < \frac{2}{3}$, this gives us a starting point from which we can produce at least one pairing with empty naughty graph, using the switching operation employed in Section 3. Thus $\nu(\emptyset) \neq \emptyset$, justifying our use of Theorem 3.9, Lemma 4.1 and Lemma 4.3.

Next we consider a few of the less common possibilities for K. The contribution to Ψ of all K with an edge of multiplicity greater than two is zero, by the definition of Ψ. The contribution of the same naughty graphs to $1/P$ is, by Lemma 4.1, at most

$$\frac{1}{6Pe(G)^{[3]}} \sum_{ii'} g_i^{[3]} \, g_{i'}^{[3]} = 0\left(\frac{g_{\max}^4}{e(G)}\right) \frac{1}{P}.$$

In a similar manner, we can dispose of those K whose underlying simple graphs have vertices of degree greater than two. The contributions to Ψ and $1/P$ are $\exp\,(0(\hat{\Delta}^2/e(G)))\Psi$ and $\exp\,(0(\hat{\Delta}^2/e(G)))/P$, respectively. Next, consider those terms of Ψ which correspond to impossible naughty graphs K (because of excessive degrees). All of these have ben counted already, except for a few cases where $1 \leq g_i \leq 3$. The contribution to Ψ here is easily seen (with the help of Lemma 4.2) to be $0(\hat{\Delta}^2/e(G))\Psi$.

We are left with the naughty graphs K for which the condition of Lemma 4.3 hold, and for which the underlying simple graph has maximum degree at most two. We must consider the possible relative errors associated with the factors $e(G)^{e(K)}/e(G)^{[e(K)]}$, $\exp(0(\hat{\Delta}e(K)/e(G)))$ and $r(K)$.

To handle the first two factors, recall the function $f(x)$ defined above. By comparing $1 + \alpha_2(ii')x^2$ with $\exp(\alpha_2(ii')x^2)$ and $1 + \alpha_1(ii')x + \alpha_2(ii')x^2$ with $\exp(\alpha_1(ii')x + \alpha_2(ii')x^2)$ we find that the coefficient of $x^{e(K)}$ in $\Psi f(x)$ is at most that of $x^{e(K)}$ in

$\exp(x\sum_{ii'\in X}\alpha_1(ii') + x^2\sum_{ii'}\alpha_2(ii'))$, which in turn is at most $\sum_{j=\lceil k/2\rceil}^{k} (\hat{\Delta}/2)^j/j!$. A straightforward argument now shows that the maximum effect of $\exp\left[0(\hat{\Delta}e(K)/e(G))\right]e(G)^{e(K)}/e(G)^{[e(K)]}$ is a factor of $\exp\left[0(\hat{\Delta}^2/e(G))\right]$.

Finally, we must investigate the factor $r(K)$. Ignoring those K we have eliminated, non-trivial contributions to $r(K)$ come from those vertices which have degree two in the underlying simple graph of K. There are three possibilities (0,1 or 2 double edges) which together provide a contribution of at most $g_i^2\hat{\Delta}^2\,\Psi/e(G)^2$ to Ψ for vertex v_i (by Lemma 4.2). The effect on $r(K)$ of this event is a factor of $1 + 0(1/g_i)$, for $g_i \neq 0$. After some routine calculations, we find that the overall effect of $r(K)$ on Ψ is a factor of $\prod_{g_i\neq 0}(1 + 0(1/g_i))^{g_i^2\hat{\Delta}^2/e(G)^2} = \exp\left[0(\hat{\Delta}^2/e(G))\right]$.

In order to estimate P we now only need to estimate Ψ.

4.5 LEMMA. If $g_{\max}^2 = 0(e(G))$, then

$$\Psi = \exp\left(\frac{1}{2e(G)^2}\sum_{i\in I} g_i^{[2]}\sum_{i\in I'} g_i^{[2]} + \frac{1}{e(G)}\sum_{ii'\in X} g_i g_{i'} + 0\left(\frac{\hat{\Delta}^2}{e(G)}\right)\right).$$

PROOF: Since $g_{\max}^2 = 0(e(G))$ we can write $1 + \alpha_2(ii') = \exp(\alpha_2(ii') + 0(\alpha_2(ii')^2)$ and $1 + \alpha_1(ii') + \alpha_2(ii')$ similarly. The rest is easy.

4.6 THEOREM. Suppose that $g_{\max} \geq 1$ and $\hat{\Delta} \leq \epsilon_1 e(G)$, where $\epsilon_1 < \frac{2}{3}$. Then

$$P(g, g', X) =$$

$$\exp\left(-\frac{1}{2e(G)^2}\sum_{i\in I} g_i^{[2]}\sum_{i\in I'} g_i^{[2]} - \frac{1}{e(G)}\sum_{ii'\in X} g_i g_{i'} + 0\left(\frac{\hat{\Delta}^2}{e(G)}\right)\right).$$

4.7 COROLLARY. Let $g = g' = (k, k, \ldots, k)$, where $1 \leq k \leq \epsilon_3 n$ for fixed $\epsilon_3 < \frac{1}{6}$. Then

$$N(g, g') = \frac{(nk)!}{(k!)^{2n}}\exp\left(-\frac{(k-1)^2}{2} + 0\left(\frac{k^3}{n}\right)\right).$$

4.8 COROLLARY. If $1 \leq k \leq o(n)$, and $g = g' = (k, k, \ldots, k)$, then

$$\log N(g, g') \sim \log\left(\frac{(nk)!}{(k!)^{2n}}\right)$$

5. Potpourri.

The results of the previous section show that the known asymptotic estimate of $N(g, g')$ for $g_{\max} = 0(\sqrt{log(n + n')})$ is in fact accurate for $g_{\max} = o(e(G)^{1/4})$. It would be of considerable interest to know if this is the limit of its validity. It is possible that the methods of this paper could be improved enough to settle this question. To begin with, Theorem 3.9 as we have it could be used to sharpen Lemmas 3.1, 3.2, 3.5 and 3.6. These in turn would imply a more accurate version of Theorem 3.9. This process could in principle be repeated, but even the first iteration would be quite complicated.

In the case $g = g' = (k, k, \ldots, k)$, we have a little experimental evidence of the accuracy of Corollary 4.7. We have computed exact values of $N(g, g')$ for $1 \le n \le 20$ and $0 \le k \le n$ [7]. A careful numerical analysis of them suggests the following possibility.

5.1 CONJECTURE. If $k \ge 1$ is fixed and $g = g' = (k, k, \ldots, k)$, then

$$n(g, g') = \frac{(nk)!}{(k!)^{2n}} \exp\left(-\frac{(k-1)^2}{2} - \frac{(k-1)^2(k^2-k+1)}{6kn} - \frac{(k-1)^5(k+1)}{12k^2n^2} + 0\left(\frac{k^5}{n^3}\right)\right).$$

If the conjecture is true, it would undoubtedly be true if k increased not too quickly with n.

Let $M_1, M_2 \in \mathbf{M}(g, g')$. We say that M_1 and M_2 are **equivalent** if M_2 can be obtained from M_1 by (independently) permuting the rows and columns. An **automorphism** of M_1 is a pair of permutations by which it is equivalent to itself. In [8] we prove that, if $3 \le g_i \le 0((n + n')^{1/2-\epsilon})$ for all i, almost no equivalence class of $\mathbf{M}(g, g')$ has members with non-trivial automorphisms. It follows, under these conditions, that the number of equivalence classes in $\mathbf{M}(g, g')$ is asymptotically $N(g, g')/n!n'!$.

Techniques similar to those of this paper can be applied to the counting of 0–1 symmetric matrices with zero diagonal and specified row sums. The results – similar to those obtained here – will be described in a later paper.

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 non-negative integer matrices with given row and column sums, *Discrete Math.* 10 (1974) 217-223.

[3] B. Bollobás, Random graphs, Proc. Eighth British Combin. Conf. (London Math. Soc. Lecture Note Series 52, Cambridge University Press, 1981) 80-102.

[4] B. Bollobás and B.D. McKay, The number of matchings in random regular graphs and bipartite graphs, to appear.

[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, Subgraphs of random graphs with specified degrees, *Congressus Numerantium* 33 (1981) 213-223.

[7] B.D. McKay, Applications of a technique for labelled enumeration, to appear.

[8] B.D. McKay and N.C. Wormald, Automorphisms of random graphs with specified degrees, to appear.

[9] M.P. Mineev and A.I. Pavlov, On the number of $(0-1)$-matrices with prescribed sums of rows and columns, *Dok. Akad. Nauk SSSR* 230, 2 (1976) 1286-1290.

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

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

[12] R.W. Robinson and N.C. Wormald, Almost all bipartite cubic graphs are hamiltonian, in volume 3.

[13] 12. N.C. Wormald, Some problems in the enumeration of labelled graphs, *Ph.D. Thesis, Dept. of Math., University of Newcastle* (1978).