Reprinted from JOURNAL OF COMBINATORIAL THEORY, Series B
Vol. 48, No. 1, February 1990
All Rights Reserved by Academic Press, New York and London
Printed in Belgium

Asymptotic Enumeration of Latin Rectangles

C. D. GODSIL*

Department of Combinatorics and Optimization, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1

AND

B. D. MCKAY†

Computer Science Department, Australian National University, GPO Box 4, Canberra, Australian Capital Territory, 2601 Australia

Communicated by the Managing Editors

Received May 21, 1986

A $k \times n$ Latin rectangle is a $k \times n$ matrix with entries from $\{1, 2, ..., n\}$ such that no entry occurs more than once in any row or column. Equivalently, it is an ordered set of k disjoint perfect matchings of $K_{n,n}$. We prove that the number of $k \times n$ Latin rectangles is asymptotically

$$(n!)^k \left(\frac{n(n-1)\cdots(n-k+1)}{n^k}\right)^n \left(1-\frac{k}{n}\right)^{-n/2} e^{-k/2}$$

as $n \to \infty$ with $k = o(n^{6/7})$. This improves substantially on previous work by Erdős and Kaplansky, Yamamoto, and Stein. We also derive an asymptotic approximation to the generalised ménage numbers, and establish a number of results on entries in random Latin rectangles. © 1990 Academic Press, Inc.

1. INTRODUCTION

A $k \times n$ *Latin rectangle* is a $k \times n$ matrix with entries from $\{1, ..., n\}$ with the property that no entry occurs more than once in any row or column. Thus an $n \times n$ Latin rectangle is nothing but a Latin square. Let $L(k, n)$ denote the number of $k \times n$ Latin rectangles. An outstanding problem is to

* Partial support for this work was provided by the NSERC under Grant A5367. This work was performed while the author was at Simon Fraser University.

† This work was performed while the author was at Vanderbilt University.

0095-8956/90 $3.00

Copyright © 1990 by Academic Press, Inc.
All rights of reproduction in any form reserved.

determine the asymptotic value of $L(k, n)$ as $n \to \infty$, with k bounded by a suitable (increasing) function of n.

The first attack on this problem was made by P. Erdős and I. Kaplansky [8], who showed that, for $k = O((\log n)^{3/2-\varepsilon})$,

$$L(k, n) \sim (n!)^k \exp\left(-\binom{k}{2}\right). \tag{1}$$

They conjectured that this result is in fact true for $k = o(n^{1/3})$; this was subsequently verified by K. Yamamoto [33]. Further progress was made by Yamamoto [36] and Stein [27], who proved that

$$L(k, n) \sim (n!)^k \exp\left(-\binom{k}{2} - \frac{k^3}{6n}\right) \tag{2}$$

for $k = O(n^{5/12-\varepsilon})$ and $k = o(n^{1/2})$, respectively.

In this paper we prove

1.1. THEOREM. *If* $0 \leqslant k = o(n^{6/7})$ *then*

$$L(k, n) \sim (n!)^k \left(\frac{[n; k]}{n^k}\right)^n \left(1 - \frac{k}{n}\right)^{-n/2} e^{-k/2}. \tag{3}$$

Here $[n; k] = n(n-1) \cdots (n-k+1)$. Theorem 1.1 is an immediate corollary to our Theorem 6.5, which is sharper but more complicated to state. We conjecture, but cannot prove, that (3) is true for $k = O(n^{1-\delta})$. As in most previous work on this topic, we estimate $L(k, n)$ by first estimating the average number of ways a randomly chosen $k \times n$ Latin rectangle can be extended to a $(k+1) \times n$ Latin rectangle by adding an extra row.

To assist the reader in understanding this paper, we now discuss the overall structure of our calculations. In Section 2 we note that to each $k \times n$ Latin rectangle R we can associate a k-regular subgraph $G = G(R)$ of the complete bipartite graph $K_{n,n}$. The number of extensions of R to a $(k+1) \times n$ Latin rectangle is equal to the number of perfect matchings of $K_{n,n}$ which contain no edge of G. However, the latter number is equal to

$$\int_0^\infty e^{-x} r(G, x)\, dx, \tag{4}$$

where $r(G, x)$ is the *rook polynomial* of G. The zeros of $r(G, x)$ are all real and lie in the interval $[0, 4k-4]$. From this it will be shown that, to evaluate (4) asymptotically for large n, we may restrict the range of integration in (4) to $[4k, \infty)$. In this range $r(G, x)$ is positive and strictly increasing.

Consequently we may write the integrand as $\exp f(x)$, where $f(x) = -x + \log r(G, x)$. In Sections 5 and 6, we will evaluate the integral by expanding $f(x)$, and then $\exp f(x)$, in a power series and integrating term by term. The coefficients in the expansion of $f(x)$ may be expressed, with some effort, as polynomials in k and the numbers of copies of various small subgraphs of G (in particular the numbers of cycles of length 4 in G). Most of this work is carried out in Section 3.

The upshot of all this is that the asymptotic number of extensions of a given $n \times k$ Latin rectangle R can be expressed in terms of n, k, and the "local structure" of $G(R)$. As we will require the *average* number of extensions of randomly chosen rectangle R, it follows that we need (amongst other things) the average number of 4-cycles to be found in $G(R)$. These calculations are performed in Section 4. With all this ground work carried out we then complete the actual asymptotic evaluation in Sections 5 and 6, as previously mentioned. Some of the power series calculations required there were too tedious to be performed by mortal hands (or at least by those belonging to the authors) and were instead done by symbolic manipulation on a computer. In fact they were done twice, on different machines.

Not much is known about the exact value of $L(k, n)$. For $k \leqslant 3$ see [26, 34] and for $k = 4$ see [1]. General formulas appear in [11] and [25], but they do not appear suitable for asymptotic analysis.

The main results of this paper were previously announced in [15].

2. Latin Rectangles and Bipartite Graphs

With each $k \times n$ Latin rectangle L we associate a k-regular bipartite graph $G = G(L)$ as follows. The vertex set of G is the union of the disjoint sets $\{c_1, ..., c_n\}$ and $\{e_1, ..., e_n\}$. A vertex c_i is adjacent to a vertex e_j iff the integer j occurs in the ith column of L. Note that, for each m $(1 \leqslant m \leqslant k)$, the edges $\{c_i, e_j\}$ such that j is in the mth row of the ith column form a perfect matching in G. Thus the k perfect matchings determined by the rows of L form a 1-factorization of G.

Conversely, given a k-regular subgraph of $K_{n,n}$ and a 1-factorization of this graph we can construct a $k \times n$ Latin rectangle. More important for our purposes is the following:

2.1. Lemma. *Let L be a $k \times n$ Latin rectangle and let $G = G(L)$. Then the number of $(k+1) \times n$ Latin rectangles $\tilde{L}$ whose first k rows coincide, in order, with the rows of L equals the number of perfect matchings $K_{n,n}$ which contain no edge of G.*

This result is easily proved given our earlier remarks. We leave its proof to the reader. The Latin rectangles $\tilde{L}$ of the lemma will be called *extensions* of L. It should be clear that we now have a graph-theoretic interpretation of the problem of counting the number of extensions of a given rectangle. If $k < n$, a Latin rectangle always has at least one extension; this is equivalent to the result that an $(n-k)$-regular bipartite graph always has a perfect matching.

The van der Waerden bound on the permanent of a doubly stochastic matrix yields the stronger conclusion that a $k \times n$ Latin rectangle has at least $n!(1-k/n)^n$ extensions. The derivation of this lower bound can be found, for example, in [6]. For a proof of the van der Waerden conjecture see [7, 10] and the exposition in [29].

3. Matchings, Walks, and Integrals

A *k-matching* in a graph G is a set of k vertex-disjoint edges. The number of k-matchings in G will be denoted by $p(G, k)$. Assume G is a subgraph of $K_{n,n}$. We define the *rook-polynomial* $r(G, x)$ by

$$r(G, x) = \sum_{k=0}^{n} (-1)^k p(G, k) x^{n-k}.$$

We adopt the convention that $p(G, 0) = 1$, so $r(G, x)$ is a monic polynomial. There is some ambiguity in our notation since if G is a subgraph of $K_{n,n}$ then it is also a subgraph of $K_{n+r,n+r}$ for $r = 1, 2, \ldots$. Unless warned otherwise the reader should always assume that n is the least integer such that $K_{n,n}$ contains G.

If G is a subgraph of $K_{n,n}$ then the relative complement G^* of G has $V(G)$ as its vertex set and $E(K_{n,n}) \setminus E(G)$ as its edge set. The next result is perhaps the fundamental tool in this paper.

3.1. Theorem. [13, Theorem 3.2]. *Let G be a subgraph of $K_{n,n}$. Then the number of perfect matchings in G^* is equal to*

$$\int_0^\infty e^{-x} r(G, x)\, dx.$$

Taken with Lemma 2.1 this result supplies us with an explicit expression for the number of extensions of a given Latin rectangle L. Although we do not know $r(G, x)$ exactly, it turns out that the above integral is relatively insensitive to the structure of G. Hence the limited information we do have allows us to compute good asymptotic estimates for the number of extensions of L.

One of the basic properties of $r(G, x)$ was first noted by Heilmann and Lieb [16].

3.2. LEMMA. *Let G be a k-regular subgraph of $K_{n,n}$. Then the zeros of $r(G, x)$ are real and lie in the half-open interval* $[0, 4k-4)$.

The fact that the zeros are real, and the bound on their value, both appear in [16, Section IV]. (We should point out that Heilmann and Lieb consider a slightly different polynomial, but their results translate readily.)

Our next result requires some further terminology. A *walk* $\mathbf{v}$ of length r in a graph G is a sequence $v_0, v_1, \ldots, v_r$ of vertices from G such that consecutive vertices are adjacent. A walk $\mathbf{v}$ of length r is *closed* if $v_0 = v_r$ and it is *reducible* if, for some i, $v_{i-1} = v_{i+1}$. In the latter case we may *reduce* $\mathbf{v}$ to a walk of length $r-2$ by omitting v_i and v_{i+1} from our sequence. This walk is closed if $\mathbf{v}$ is. Of course, a walk which is not reducible will be called *irreducible*.

Given any walk $\mathbf{v}$ we may, by a sequence of reductions, obtain an irreducible walk $\mathbf{v}'$. The walk $\mathbf{v}'$ is uniquely determined by $\mathbf{v}$. (This is equivalent to a standard result concerning free groups and is also proved in detail in [12].) It is quite possible that $\mathbf{v}'$ is a single vertex, in which case $\mathbf{v}$ is said to be *totally reducible.*

Finally, a walk $\mathbf{v} = (v_0, \ldots, v_r)$ is *tree-like* if, for each $i = 0, \ldots, r$, the walk $(v_0, \ldots, v_i)$ reduces to a path, i.e., to a walk where all vertices involved are distinct. We denote by w_r half the number of closed tree-like walks with length $2r$ in G. Note that a closed tree-like walk is totally reducible and so must have even length. We have:

3.3. LEMMA [12, Theorem 3.6(b)]. *Let G be a subgraph of $K_{n,n}$. Then*

$$\sum_{r=0}^{\infty} w_r x^{-1-r} = r'(G, x)/r(G, x)$$

and so $w_r = \sum_{i=1}^{n} \lambda_i^r$, *where* $\lambda_1, \ldots, \lambda_n$ *are the zeros of* $r(G, x)$.

The explicit expression for w_r follows from the formal power series identity by partial fractions. Hence the power series converges if $|x| > \max\{\lambda_i\}$.

In particular, it is a valid expansion for $r'(G, x)/r(G, x)$ when $x \geqslant 4k-4$ and G is a k-regular subgraph of $K_{n,n}$.

If G is a k-regular graph, then the number of totally reducible walks of length $2r$ which start at a given vertex equals the total number of closed walks of length $2r$ starting at a given vertex in the infinite tree with each vertex of valency k. This leads to the following:

3.4 LEMMA [18, 19]. *Let G be a k-regular graph with n vertices and let*

u_r be the number of totally reducible walks of length $2r$ starting at a given vertex v in G. Then

(a) *the number of totally reducible closed walks of length $2r$ in G is nu_r,*

(b) $u_r = \sum_{j=0}^{r} \binom{2r}{j} \frac{2r-2j+1}{2r-j+1} (k-1)^j$,

(c) $\sum_{r=0}^{\infty} u_r x^r = 2(k-1)/[k-2+k(1-4(k-1)x)^{1/2}]$ *for* $|x| < (4(k-1))^{-1}$, *and*

(d) $\sum_{r=1}^{\infty} (u_r/r) x^{-r} = \log [\eta^2((k-\eta)/(k-1))^{k-2}]$ *for* $|x| > 4(k-1)$, *where* $2\eta(k-1) = x - (x^2 - 4(k-1)x)^{1/2}$.

The importance of this lemma lies in the fact that it implies that the number of totally reducible closed walks of length $2r$ in a k-regular graph G is determined only by n and k. This will prove extremely useful because nu_r is both an upper bound for and, for small r, a reasonable first approximation to w_r.

3.5. LEMMA. *Let G be a k-regular bipartite graph with $2n$ vertices and let $\varepsilon_r = nu_r - w_r$. Then*

(a) $\varepsilon_r = 0$ $(r = 1, 2, 3)$,

(b) $\varepsilon_4 = 4s$,

(c) $\varepsilon_5 = 40(k-1)s$,

(d) $\varepsilon_6 = 264(k-1)^2 s + 6h - 24b$, *and*

(e) $\varepsilon_m = O(nk^{m-1})$ *as $n \to \infty$ with m fixed.*

Here s is the number of squares (i.e., cycles of length 4) in G, h is the number of hexagons, and b is the number of copies of $K_{2,3}$.

Proof. Clearly ε_r equals half the number of totally reducible walks of length $2r$ which are not tree-like. The subgraph induced by the vertices used in such a walk must contain a cycle, since every closed walk in a tree is tree-like. (This is easy to prove and is spelled out as Lemma 3.4 in [12].) Thus $\varepsilon_r = 0$ for $r = 1, 2$, and 3. Since each square provides eight totally reducible walks of length 8 that are not tree-like, $\varepsilon_4 = 4s$.

It may appear that ε_5 should depend on the number of "squares with one end-vertex added" as well as on s. However, the number of the former is determined by s and k. Similarly arguments show that ε_6 depends only on

k, s, h, and b. For both these cases ((c) and (d)) we determined the relevant coefficients by computer.

Finally, we prove (e). The claim is true for $m \leqslant 3$, so fix $m \geqslant 4$. For any walk W counted by ε_m, let $G(W)$ be the subgraph of G induced by W. $G(W)$ is clearly connected and, since W is ~~not~~ totally reducible, it has at most m vertices. However, there are only $O(nk^{m-1})$ connected induced subgraphs of order m or less in G and, since m is bounded, each contributes $O(1)$ to ε_m.

We note that one of the reasons for the success of our approach is the use of the integral formula in Theorem 3.1 in place of inclusion–exclusion. This integral formula is also obtained by Joni and Rota in [17]. Related results appear in [2, 9].

4. Random $k \times n$ Latin Rectangles

In this section we estimate the average number of copies of a given graph contained in a graph $G(R)$, where R is a randomly chosen $k \times n$ Latin rectangle. (Here all $k \times n$ rectangles are equally probable.) We will need some further terminology before we can proceed.

Assume k and n are given. By a *triple* we mean an ordered triple $(i, j; \alpha)$ such that $1 \leqslant i \leqslant k$ and $1 \leqslant j$, $\alpha \leqslant n$. Two triples with the same first coordinate will be said to lie in the same *row*; the second coordinate similarly determines the *column* of the triple. If $x = (i, j; \alpha)$ is a triple we will refer to the ordered pair (i, j) as the *position* of x and to α as the *contents* of x. We denote the former by $\operatorname{pos}(x)$ and the latter by $\operatorname{cont}(x)$. A set L of triples is *Latin* if no two triples agree in more than one coordinate. The number of triples in row i of L will be denoted by $r_i(L)$ and the maximum value of $r_i(L)$ by $r(L)$.

Given the terminology just defined, we can view a $k \times n$ Latin rectangle as a Latin set of kn triples. If L is a Latin set of triples and $H \subseteq L$ then $N(L, H)$ denotes the number of $k \times n$ Latin rectangles R such that $R \cap L = H$. Of course we are mostly interested in estimating $N(L, L)$, but the quantities $N(L, H)$ will arise in the course of our calculations. Finally, we shall use $[a; k]$ to denote the falling factorial $a(a-1) \cdots (a-k+1)$.

The following result is the main tool in this section.

4.1. Theorem. *Let k and n be given. Let L be a Latin set of triples, let H be a subset of L, and let J be a subset of $L \backslash H$. Then we have*:

(a) *If* $r(L) < n - 2k$, $N(L, H)/N(L, H \cup J) \geqslant \prod_{i=1}^{k} [n - 2k + 1 - r_i(H); r_i(J)]$, *and*

(b) *if* $r(L) < n-4k$, $N(L, H)/N(L, H\cup J) \leqslant (1-\beta(n,k))^{-|J|}$ $\prod_{i=1}^{k} [n-1-r_i(H);\, r_i(J)]$,
where $\beta(n,k) = (2k-1)(n-2k+1-r(H\cup J))^{-1}$.

We remark that it will follow from our proof that if $r(L) \leqslant n-4k$ then $N(L, H)$ and $N(L, H\cup J)$ are both non-zero. The bulk of this proof will be presented as two separate lemmas. These require some further notation.

Call two $k\times n$ Latin rectangles R_1 and R_2 *related* if one can be obtained from the other by interchanging the contents of two triples in the same row. Let x be a triple in $L\backslash H$ and let M be the number of pairs of $k\times n$ rectangles (R_1, R_2) such that R_1 and R_2 are related and

$$R_1\cap L = H,\; R_2\cap L = H\cup\{x\}.$$

4.2. LEMMA. *With notation as above, if x is in row i and $r_i(H)\leqslant n-1$ we have* $N(L, H\cup\{x\})(n-2k+1-r_i(H)) \leqslant M \leqslant N(L, H\cup\{x\})(n-r_i(H)-1)$.

Proof. Choose a random $k\times n$ Latin rectangle R_2 such that $R_2\cap L = H\cup\{x\}$. To make a related rectangle R_1 such that $R_1\cap L = H$, we choose a triple y in $R_2\backslash(H\cup\{x\})$ in the same row as x and interchange its contents with those of x. There are at most $n-r_i(H)-1$ choices for y, which yields the upper bound of the lemma. However if the contents of the triple chosen as y coincide with the contents of a triple of R_2 in the same column as x, or if the contents of x coincide with the contents of some triple in the same column as y, our interchange will produce a non-Latin set of triples. These constraints eliminate at most $2(k-1)$ possible choices of y. This still leaves at least $n-2k+1-r_i(H)$ possibilities, which implies the lower bound stated.

4.3. LEMMA. *With notation as before, if $r_i(H) < n-2k$ we have*

$$N(L, H)(1-(2k-1)(n-2k-r(H))^{-1}) \leqslant M \leqslant N(L, H).$$

Proof. Let R_1 be a $k\times n$ Latin rectangle chosen at random subject to the condition $R_1\cap L = H$. Let x_1 be the triple in R_1 with $\mathrm{pos}(x_1) = \mathrm{pos}(x)$. Let y be the unique triple in R_1 in the same row as x (and x_1) with $\mathrm{cont}(y) = \mathrm{cont}(x)$. If we interchange the contents of x_1 and y then we obtain either

(1) a Latin rectangle R_2 with $R_2\cap L = H\cup\{x\}$,

(2) a Latin rectangle R_2 with $R_2\cap L = H\cup\{x, y\}$ for some y', or

(3) a non-Latin set of triples.

Since the triple y is unique (being determined by its contents) we have

$M \leqslant N(L, H)$. To obtain a lower bound on M we need to estimate the probability that (2) or (3) occurs. We will begin with (2).

Note first that since, as just established, $M \leqslant N(L, H)$ it follows, using the lower bound in Lemma 4.2, that

$$\frac{N(L, H \cup \{x\})}{N(L, H)} \leqslant (n-2k+1-r(H))^{-1}.$$

This holds for all L, H satisfying the given conditions and so can also be applied wth $H \cup \{x\}$ in place of H and y in place of x to obtain a bound on $N(L, H \cup \{x, y\})/N(L, H \cup \{x\})$. Multiplying these two inequalities together yields that $N(L, H \cup \{x, y\})/N(L, H)$ is at most $[n-2k+1-r(H); 2]^{-1}$, which is never greater than $(n-2k+1-r(H))^{-1}$.

Obtaining a bound in case (3) will cause us considerably more difficulty. First, (3) can occur for two distinct reasons;

(3a) Some triple of R_1 in the same column as x_1 already contains the contents required in x.

(3b) Some triple of R_1 in the same column as y contains the contents of x_1.

(Thus (3a) arises if interchanging the contents of x_1 and y yields two triples in the x_1-column with the same contents; (3b) arises when the clash occurs in the y-column.) We will first bound the probability that (3a) occurs. Assume that x_2 is a triple in the same column as x_1 with $\text{cont}(x_2) = \text{cont}(x)$. (Note that $x_2 \notin H$, since H is Latin.)

Let A denote the set of all $k \times n$ rectangles R such that $R \cap L = H$ and $x_2 \in R$. Let B denote the set of those rectangles R such that $R \cap L = H$ and $x_2 \notin R$. (Thus $|A \cup B| = N(L, H)$.)

If $R_1 \in B$ then there is a unique triple z in the same row as x_2 with $\text{cont}(z) = \text{cont}(x)$. Hence there is at most one related rectangle R_2 such that $R_2 \in A$. Thus the number of pairs (R_1, R_2) of related rectangles with R_1 in B and R_2 in A is at most $|B|$.

Suppose conversely that we are given a rectangle R_2 in A. Choosing a triple z in the x_2-row and swapping its contents with those of x_2 gives a rectangle R_1 in B unless

(a) $\text{pos}(z) = \text{pos}(x_2)$,

(b) after the swap, the triple z' of R_1 with position $\text{pos}(z)$ belongs to L,

(c) after the swap, the triple x_2' of R_1 with position $\text{pos}(x_2)$ belongs to L,

(d) $z \in H$, or

(e) the swapping produces a non-Latin set of triples.

Conditions (a), (b), and (c) each exclude at most one choice of z in the x_2-row. Condition (d) excludes at most $r(H)$ choices and (e) excludes at most $2(k-1)$. This leaves at least $n-2k-1-r(H)$ choices. Hence the number of related pairs (R_1, R_2) with R_1 in B and R_2 in A is at least $(n-2k-1-r(H))\,|A|$. Combining this with our upper bound of $|B|$ on the same number, and recalling that $|A|+|B|=N(L,H)$, we obtain $|A| \leqslant N(L,H)(n-2k-r(H))^{-1}$.

We thus have a bound on the probability that x_2 contains cont(x). The probability that some triple in the x-column contains cont(x) is at most $k-1$ times this bound — i.e., the probability that (3a) occurs is at most $(k-1)/(n-2k-1-r(H))$.

To bound the probability that (3a) occurs, note that (3a) and (3b) are dual under the duality induced by interchanging the second and third coordinates of all triples. Thus, (3b) occurs with probability at most $(k-1)/(n-2k-1-r(H))$. (We are indebted to the referee for this argument.)

In summary, the probability that case (2) or case (3) occurs is at most $(2k-1)(n-2k-r(H))^{-1}$ and this suffices to complete the proof of the lemma.

It follows from Lemma 4.3 that if $N(L,H)>0$ and $r(H)<n-4k$ then $M>0$. Hence $N(L, H\cup\{x\})>0$. In particular if $r(L)\leqslant n-4k$, we deduce that $N(L,H)>0$ for all subsets H of L. This justifies the claims made immediately following the statement of Theorem 4.1.

4.4. *Proof of Theorem* 4.1. Combining the upper bound from Lemma 4.3 and the lower bound from Lemma 4.2 we obtain

4.5. $N(L,H)/N(L,H\cup\{x\}) \geqslant (n-2k+1-r_i(H))$.

The lower bound from Lemma 4.3 and the upper bound from Lemma 4.2 together imply that

4.6. $N(L,H)/N(L,H\cup\{x\}) \leqslant (1-(2k-1)(n-2k-r(H))^{-1})^{-1}(n-1-r_i(H))$.

Theorem 4.1 follows from 4.5 and 4.6 by a trivial induction argument.

4.7. THEOREM. *Let L be a Latin set of triples such that for some $v>4$, $r(L)\leqslant n-vk$. Then the probability $P(k,n,L)$ that a random $k\times n$ Latin rectangle contains L is*

$$n^{-|L|}\exp(O(k|L|(n-2k+1-r(L))^{-1})) \qquad \text{as} \quad n\to\infty.$$

Proof. We have

$$P(k, n, L) = N(L, L) \Big/ \sum_{S \subseteq L} N(L, S)$$
$$= N(L, L)/(N(L, \varnothing) \sum_{S \subseteq L} (N(L, S)/N(L, \varnothing))).$$

Theorem 4.1 supplies bounds on $N(L, S)/N(L, \varnothing)$. Since there are $\binom{|L|}{r}$ subsets S of L with $|S| = r$ we obtain

$$1 \leqslant \sum_{S \subseteq L} N(L, S)/N(L, \varnothing) \leqslant \sum_{r=0}^{|L|} \binom{|L|}{r} (n - 2k + 1 - r(L))^{-r}$$
$$= (1 + (n - 2k + 1 - r(L))^{-1})^{|L|}$$
$$\leqslant \exp(|L|(n - 2k + 1 - r(L))^{-1}).$$

Applying Theorem 4.1 to $N(L, \varnothing)/N(L, L)$ yields

$$\prod_{i=1}^{k} [n - 2k + 1; r_i(L)] \leqslant \frac{N(L, \varnothing)}{N(L, L)} \leqslant \left(1 - \frac{2k-1}{\gamma}\right)^{-|L|} \prod_{i=1}^{k} [n - 1; r_i(L)],$$

where $\gamma = n - 2k + 1 - r(L)$. Comparing these two pairs of bounds yields the assertion of the theorem.

Our main application of Theorem 4.7 will be to estimate the average number of subgraphs of specified type in $G(R)$, where R is a randomly chosen $k \times n$ Latin rectangle. If H is a bipartite graph, let $n(R, H)$ denote the number of subgraphs K of $G(R)$ isomorphic to H, where the isomorphism must map "row" vertices of K to vertices in H of the first colour and "column" vertices to vertices in H of the second colour.

4.8. THEOREM. *Let H be a fixed bipartite graph with n_1 vertices of the first colour, n_2 of the second colour, and m edges. Suppose that the group of automorphisms of H preserving the colours has order a and that H can be properly edge-coloured with k colours in $\psi(k)$ ways. Then, if $k \leqslant n/5$, the expected value of $n(R, H)$ is*

$$n^{n_1 + n_2 - m} \psi(k) a^{-1} \left(1 + O\left(\frac{k}{n}\right)\right) \qquad as \quad n \to \infty.$$

Proof. The number of bipartition-preserving bijections from $V(H)$ into $V(G(R))$ is $[n; n_1][n; n_2]/a$, which equals $n^{n_1 + n_2} a^{-1}(1 + o(n^{-1}))$ for fixed H. The number of ways of assigning the edges of H to the rows of R is $\psi(k)$. The probability that the image of one of these bijections is isomorphic to H is, by Theorem 4.7, equal to $n^{-m}(1+O(k/n))$. This yields our claim immediately.

As a typical application, we find that the expected number of cycles of length r in $G(R)$ is

$$n(R, C_r)=((k-1)^r+k-1)\,r^{-1}\left(1+O\left(\frac{k}{n}\right)\right)$$

if r is even. (For the number of edge-colourings of a cycle see, e.g, [3, p. 59].) Interestingly, this value is different from the value for random k-regular bipartite graphs, which is asymptotically $(k-1)^r/r$. (See [20, 21, 31].) When H is a cycle we can establish more information about the distribution of $n(R, H)$.

4.9. THEOREM. *Let $r \geqslant 4$ and $t \geqslant 0$ be fixed integers, with r even. Then the tth factorial moment of $n(R, C_r)$ is*

$$n_t(R, C_r)=(((k-1)^r+k-1)/r)^t\,(1+O(k/n))$$

as $n \to \infty$ with $k \leqslant n/5$.

Proof. By its definition, $n_t(R, C_r)$ is the expected number of ordered t-tuples of distinct r-cycles in $G(R)$. Consider first those t-tuples where the cycles are vertex-disjoint. By Theorem 4.8, the expected number of these is $(((k-1)^r+k-1)/r)^t\,(1+O(k/n))$. Next consider the cases where the t r-cycles form a subgraph H' with $s \leqslant rt-1$ vertices. The number of edge colourings of such a graph with k colours is no greater than the number of such colourings of t vertex disjoint copies of C. Also, the automorphism group of H' is clearly at most $r^t t!$ times smaller than that of tC_r. Since r and t are constant there are a bounded number of isomorphism types for H', and so these cases contribute a factor of at most $1+O(1/n)$.

It is worth noting that $((k-1)^r+(k-1)/r)^t$ is the tth factorial moment of the Poisson distribution with mean $((k-1)^r+k-1)/r$. Thus, if $k=O(n)$, we find that under suitable normalization the distribution of $n(R, C_r)$ converges to this Poisson distribution. If k is fixed then the convergence is pointwise and we can infer, for example, that $G(R)$ is free of r-cycles with probability approaching $\exp(-((k-1)^r+k-1)/r)$. This is easily extendible to the study of the girth of $G(R)$ and other such things.

We next find a more accurate estimate of $n_1(R, C_4)$, the expected value of $n(R, C_4)$. The increased accuracy will become important in Section 6.

4.10. THEOREM. *If $k \leqslant n/5$ then, as $n \to \infty$,*

$$n_1(R, C_4)=\frac{k(k-1)(k^2-3k+3)}{4}+\frac{k(k-1)(k-2)^2}{2n}+O\left(\frac{k^6}{n^2}\right).$$

Proof. The method we use is essentially that of the early part of this section. A 4-cycle in $G(R)$ can appear in R in one of three ways, depending on whether it occupies two, three, or four rows. We will treat each of these cases separately, but only treat the first in detail. Our principal tool will be the following weak amalgam of parts (a) and (b) of Theorem 4.1:

$$\text{If } k \leqslant n/5 \text{ and } r(L) = o(1) \text{ then } N(L, H \cup \{x\})/N(L, H) = \frac{1}{n}\left(1 + O\left(\frac{k}{n}\right)\right). \tag{1}$$

Case (i). Suppose the 4-cycle occupies exactly two rows. There are $n(n-1)\binom{n}{2}\binom{k}{2}$ possibilities, of which a representative is

$$\begin{bmatrix} 1 & 2 & & \\ 2 & 1 & & \\ & & & \\ & & & \end{bmatrix}.$$

We seek the probability that this particular 4-cycle occurs. Let s_1 be the set of $k \times n$ Latin rectangles which contain this 4-cycle and let s_2 be the set of those which contain instead

$$\begin{bmatrix} 1 & 2 & & \\ 2 & x & & \\ & & & \\ & & & \end{bmatrix},$$

where $x \notin \{1, 2\}$. Let M be the number of related pairs (R_1, R_2) such that $R_1 \in S_1$ and $R_2 \in S_2$.

The probability that a random rectangle is in $S_1 \cup S_2$ is exactly $1/(n(n-1)^2)$. Now choose a random R_1 from S_1. To generate a related R_2 in S_2 we need to select a position in row 2 (but not in the first two columns) and interchange its contents with those of the (2, 2) position. Of the $n-2$ possible positions, $k-2$ have contents the same as some position in column 2 and $k-2$ are in the same column as a 1. By (1), $O(k^2/n)$ have both problems at once. Thus

$$M = |S_1|(n - 2k + 2 + O(k^2/n)).$$

Next, choose a random rectangle R_2 from S_2. This will have a related companion in S_1 unless column 2 contains a 1 (probability $(k-2)/n + O(k^2/n^2)$), or the column in row 2 which contains a 1 also contains an entry the same as that in the (2, 2) position of R_2 (probability $(k-1)/n+O(k^2/n^2)$), or both problems occur at once (probability $O(k^2/n^2)$). Then

$$M=|S_2|(1-(2k-3)/n+O(k^2/n^2)).$$

Comparing the two estimates of M, we find

$$|S_1|/(|S_1|+|S_2|)=\frac{1}{n}+O\left(\frac{k^2}{n^3}\right).$$

Therefore the expected number of C_4's of this type is $\frac{1}{4}k(k-1)(1+O(k^2/n^2))$.

Case (ii). Suppose the 4-cycle occupies exactly three rows. There are $n^2(n-1)^2\,k(k-1)(k-2)/2$ possibilities, each of which occurs with probability $(1+1/n+O(k^2/n^2))/n^2(n-1)^2$. Thus the expected number of 4-cycles of this type is $\frac{1}{2}k(k-1)(k-2)(1+1/n+O(k^2/n^2))$.

Case (iii). Suppose that the 4-cycle occupies four rows. Then there are $n^2(n-1)^2\,k(k-1)(k-2)(k-3)/4$ possibilities, each of which occurs with probability $(1+2/n+O(k^2/n^2))/n^2(n-1)^2$. Hence the expected number is

$$\tfrac{1}{4}k(k-1)(k-2)(k-3)(1+2/n+O(k^2/n^2)).$$

Adding these three estimates yields the theorem.

5. First calculations

Suppose $G=G(L)$, where L is a $k\times n$ Latin rectangle. Then the number of extensions of L is, by Theorem 3.1, equal to

$$E_k=\int_0^\infty e^{-x}r(x)\,dx,$$

where we abbreviate $r(G,x)$ to $r(x)$.

Let

$$I_k=\int_0^{4k} e^{-x}r(x)\,dx,\qquad J_k=\int_{4k}^\infty e^{-x}r(x)\,dx.$$

In this section we show that asymptotically I_k is negligible compared to J_k. This means that we may concentrate on evaluating J_k rather than E_k. We then show that J_k can be estimated sufficiently accurately by integrating over a suitable finite interval, rather than over $[4k,\infty)$.

5.1. LEMMA. *If G is a k-regular bipartite graph on $2n$ vertices then $|I_k| \leqslant \int_0^{4k} e^{-x}(k+x/2)^n\,dx$.*

Proof. Suppose $r(x) = \prod_{i=1}^{n}(x-\lambda_i)$. By the inequality for arithmetic and geometric means,

$$|r(x)|^{1/n} \leqslant \frac{1}{n}\sum_{i=1}^{n} |x-\lambda_i|. \tag{1}$$

Let $b(x) = \sum_{i=1}^{n} |x-\lambda_i|$. We claim that $b(x) \leqslant n(k+x/2)$ for $0 \leqslant x \leqslant 4k$.

The λ_i satisfy the contraints

(a) $0 \leqslant \lambda_i \leqslant 4k$ and

(b) $\sum_{i=1}^{n} \lambda_i = nk$,

where the latter holds because $nk = w_1$.

Consider choosing $\{\lambda_1, ..., \lambda_n\}$ to maximise $b(x)$ for fixed $x \in [0, 4k]$. If $0 < \lambda_i \leqslant x$ and $x \leqslant \lambda_j < 4k$ $(i \neq j)$, we can increase $b(x)$ while still satisfying (a) and (b) if we replace λ_i by $\lambda_i - m$ and λ_j by$\lambda_j + m$, where $m = \min\{\lambda_i, 4k-\lambda_j\}$. Thus, the sets $\{\lambda_1, ..., \lambda_n\}$ which maximise $b(x)$ either have t of the λ_i equal to 0 and the rest in $[x, 4k]$ or have t of the λ_i in $[0, x]$ and the rest equal to $4k$, for some t. In each case, we must have $t \leqslant 3n/4$ if (b) is satisfied, which easily leads to the conclusion that $b(x) \leqslant n(k+x/2)$.

5.2. LEMMA. *If G is a k-regular bipartite graph on $2n$ vertices and $k \leqslant n/10$ then $|I_k|/E_k < 0.912^n$.*

Proof. First we have, by Lemma 5.1,

$$\begin{aligned} |I_k| &\leqslant \int_0^{4k} e^{-x}(k+x/2)^n\,dx \\ &= 2^{-n}e^{2k}\int_{2k}^{6k} e^{-u}u^n du \\ &\leqslant 2^{-n}e^{2k}n!. \end{aligned}$$

On the other hand each zero of $r(x)$ lies in the interval $[0, 4k)$ and so $r(x) \geqslant (x-4k)^n$ for all $x \geqslant 4k$. Hence

$$J_k = \int_{4k}^{\infty} e^{-x}r(x)\,dx \geqslant \int_{4k}^{\infty} e^{-x}(x-4k)^n\,dx \tag{2}$$

$$= e^{-4k}n!. \tag{3}$$

Therefore $|I_k|/J_k \leqslant 2^{-n}e^{6k}$. Setting $k = n/10$ here yields our result.

The previous result can be strengthened by appealing to the van der Waerden bound. For as we remarked in Section 2 this bound implies that $E_k \geqslant n!(1-k/n)^n$. Thus we obtain

$$|I_k|/E_k \leqslant 2^{-n} e^{2k} \left(1 - \frac{k}{n}\right)^{-n}.$$

If $k \leqslant n/5$ then the RHS of this inequality is less than 0.933^n. In summary, by appealing to the van der Waerden bound, our restriction $k \leqslant n/10$ in Lemma 5.2 can be weakened to $k \leqslant n/5$. (The exact value of the constant is unimportant, of course, provided it is less than one.)

Let ψ be the value of x in the interval $[4k, \infty)$ which maximizes $e^{-x}r(x)$. (We will see in the proof of Lemma 5.3 that this value is unique.) Define

$$J_k(a) = \int_{\psi - a}^{\psi + a} e^{-x} r(x)\, dx.$$

5.3. LEMMA. *Let G be a k-regular bipartite graph with n vertices. Assume $k \leqslant n/5$. Then for any ε such that $0 < \varepsilon < \frac{1}{6}$, we have uniformly $J_k(n^{1/2+\varepsilon})/J_k = 1 - O(n \exp(-n^{2\varepsilon}))$.*

Proof. Let the zeros of $r(x)$ be $\lambda_1, \lambda_2, \ldots, \lambda_n$ in non-increasing order. Let $f(x) = e^{-x} r(x)$. It is easily checked that $f'(x)$ vanishes precisely when $r'(x)/r(x) = 1$. Using partial fractions we see this is equivalent to

$$\sum_{i=1}^{n} \frac{1}{x - \lambda_i} = 1. \tag{4}$$

Clearly for $x > \lambda_1, r'(x)/r(x)$ is a decreasing function of x. Hence $f(x)$ has a unique maximum in the range $[4k, \infty)$, as we claimed above. Since the λ_i are all non-negative it follows from (4) that $n \leqslant \psi < n + 4k$.

To determine the relative difference between J_k and $J_k(a)$ we will consider separately the contribution of the intervals $[4k, \psi - n^{1/2+\varepsilon}]$ and $[\psi + n^{1/2+\varepsilon}, \infty)$ to J_k. We begin with the lower tail.

We have

$$\begin{aligned} f(\psi + t)/f(\psi) &= e^{-t} \prod_{i=1}^{n} \left(1 + \frac{t}{\psi - \lambda_i}\right) \\ &= \exp\left(-t + \sum_{i=1}^{n} \log\left(1 + \frac{t}{\psi - \lambda_i}\right)\right). \end{aligned}$$

Assuming $|t/(\psi-\lambda_i)| \leqslant c < 1$ for some constant c, we can expand the logarithmic terms here in power series to yield uniformly

$$f(\psi+t)/f(\psi)$$
$$= \exp\left(-t + t\sum \frac{1}{\psi-\lambda_i} - (t^2/2)\sum \frac{1}{(\psi-\lambda_i)^2} + O\left(\sum \frac{t^3}{(\psi-\lambda_i)^3}\right)\right). \tag{5}$$

From (4) we have $\sum(\psi-\lambda_i)^{-1} = 1$. Consequently

$$\frac{1}{n}\sum \frac{1}{(\psi-\lambda_i)^2} \geqslant \left(\frac{1}{n}\sum \frac{1}{\psi-\lambda_i}\right)^2 = n^{-2}$$

and so $\sum(\psi-\lambda_i)^{-2} \geqslant n^{-1}$. Since $\lambda_i < 4k$, $\psi \gg n$ and $k \leqslant n/5$ we also have $\sum(\psi-\lambda_i)^{-3} = O(n^{-2})$. Applying all this to (5) we find that

$$f(\psi+t)/f(\psi) \leqslant \exp\left(-\frac{t^2}{2n} + O\left(\frac{t^3}{n^2}\right)\right). \tag{6}$$

Now, since $e^{-x}r(x)$ has just one turning point in the interval $[4k, \infty)$ and is a decreasing function of x for large enough x, $f(\psi+t)/f(\psi)$ is bounded above on $[4k, \psi - n^{1/2+\varepsilon}]$ by its value at $\psi - n^{1/2+\varepsilon}$. Therefore the contribution of this interval to J_k is bounded above by

$$(\psi - 4k - n^{1/2+\varepsilon}) \exp(-\tfrac{1}{2}n^{2\varepsilon} + O(n^{3\varepsilon-1/2})). \tag{7}$$

We now must consider the upper tail of our integral. Set

$$g(t) = -t + \sum_{i=1}^{n} \log(1 + t/(\psi-\lambda_i)).$$

Then

$$g'(t) = -1 + \sum_{i=1}^{n} ((\psi-\lambda_i)(1 + t/(\psi-\lambda_i)))^{-1}.$$

Note that $\exp g(t) = f(\psi+t)/f(\psi)$. Clearly $g'(t)$ decreases as t increases. For $t \geqslant a$ this means we can bound $g(t)$ above by its tangent at $t = a$. We have

$$\begin{aligned} g'(t) &\leqslant -1 + \sum ((\psi-\lambda_i)(1+t/\psi))^{-1} \\ &= -1 + (1+t/\psi)^{-1}\sum(\psi-\lambda_i)^{-1} \\ &= -1 + (1+t/\psi)^{-1} && (8') \\ &= -t/\psi + O(t^2/\psi^2) && (8) \end{aligned}$$

uniformly for $t \geqslant 0$. Thus $g'(n^{1/2+\varepsilon}) \leqslant -n^{1/2+\varepsilon}/\psi + O(n^{1+2\varepsilon}/\psi^2)$.

It follows that, if $a = n^{1/2+\varepsilon}$,

$$\int_a^\infty \exp g(t)\, dt \leqslant \int_a^\infty \exp(g(a) + (t-a)\, g'(a))\, dt$$

$$= \exp g(a) \int_0^\infty \exp(sg'(a))\, ds$$

$$= -\exp(g(a))/g'(a).$$

~~An argument similar to the one used to obtain~~ (8) yields $g'(t) \leqslant -t/(\psi + \lambda_1)$. Therefore $-\exp(g(a))/g'(a)$ is bounded above by

$$\frac{\psi + \lambda_1}{a} \exp\left(-a + \sum_{i=1}^{n} \log\left(1 + \frac{a}{\psi - \lambda_i}\right)\right)$$

$$= \frac{\psi + \lambda_1}{a} \exp\left(-a + \sum \frac{a}{\psi - \lambda_i} - \frac{1}{2} \sum \frac{a^2}{(\psi - \lambda_i)^2} + O\left(\sum \frac{a^3}{(\psi - \lambda_i)^3}\right)\right)$$

$$\leqslant \frac{\psi + \lambda_1}{n^{1/2+\varepsilon}} \exp\left(-\frac{1}{2} n^{2\varepsilon} + O(n^{-1/2+3\varepsilon})\right). \tag{9}$$

As $\psi < n + 4k$ and $k \leqslant n/5$, $(\psi + \lambda_1)/n^{1/2+\varepsilon} = O(n^{1/2-\varepsilon})$. Given this, (7) and (9) together yield the lemma. (Note that since $\varepsilon < \frac{1}{6}$, $O(n^{-1/2+3\varepsilon}) = O(1)$, and this term is negligible.)

6. Computing the Numbers of Extensions and Latin Rectangles

Throughout this section we will assume that $1 \leqslant k = O(n^{1-\delta})$ for some fixed $\delta > 0$. Thus, by Lemma 5.3, we can restrict our attention to the interval $[\psi - n^{1/2+\varepsilon}, \psi + n^{1/2+\varepsilon}]$, where $\varepsilon > 0$ and ψ is the point where $e^{-x}r(x)$ is greatest. We will estimate the integral by approximating $e^{-x}r(x)$ in this interval. The first step will be to estimate ψ. In the following lemma and throughout this section, each $O(\cdot)$ is for $n \to \infty$, and the implied constant is independent of k.

6.1. Lemma. *Let* $1 \leqslant k \leqslant n^{1-\delta}$ *and define* θ *to be the value of* x *at which* $\phi(x) = e^{-x}x^n \exp(-n \sum_{r=1}^\infty u_r/(rx^r))$ *is greatest. Then*

(i) $(k-1)\, n^2/\theta^2 + (k^2/n - k + 2)\, n/\theta - 1 = 0$

(ii) $\dfrac{n}{\theta} = 1 - \dfrac{k}{n} + \dfrac{k}{n^2} + \dfrac{k(k-2)}{n^3} + O\left(\dfrac{k^3}{n^4}\right)$ *and*

(iii) $\psi = \theta + O(n^{1/2})$.

Proof. Part (i) follows from the value of the generating function $u(x)$, as given in Lemma 3.4. This can be solved to give (ii).

Let $f(x)=e^{-x}r(x)$. From the proof of Lemma 5.3 we see that

$$\frac{f(\psi+t)}{f(\psi)}\leqslant\begin{cases}\exp\left(-\frac{t^2}{2}\sum\frac{1}{(\psi-\lambda_i)^2}+O(t^3/n^2)\right), & \text{if } |t|\leqslant n^{1/2+\varepsilon},\\ \exp(-\frac{1}{2}n^{2\varepsilon}+O(n^{-1/2+3\varepsilon})), & \text{if } |t|\geqslant n^{1/2+\varepsilon}.\end{cases}$$

Let $x\in[\psi-n^{1/2+\varepsilon},\psi+n^{1/2+\varepsilon}]$. Then $f(x)/\phi(x)=\exp(\sum_{r=4}^{\infty}\varepsilon_r/rx^r)$ by Lemma 3.3. Now $\varepsilon_r\leqslant n(4k)^r$ generally and $\varepsilon_r=O(nk^{r-1})$ for fixed r by Lemma 3.5. Define $N=6+\lceil 1/\delta\rceil$. Then

$$\begin{aligned}1\leqslant\frac{f(x)}{\phi(x)}&\leqslant\exp\left(\sum_{r=4}^{N-1}\frac{\varepsilon_r}{rx^r}+\sum_{r=N}^{\infty}\frac{\varepsilon_r}{rx^r}\right)\\&\leqslant\exp\left(\sum_{r=4}^{N-1}\frac{O(nk^{r-1})}{rx^r}+\sum_{r=N}^{\infty}\frac{n(4k)^r}{rx^r}\right)\\&=\exp(o(1)),\qquad\text{since}\quad k=O(n^{1-\delta})\quad\text{and}\quad\psi\geqslant n.\end{aligned}$$

Therefore f must have its maximum within $O(n^{1/2})$ of that of ϕ.

The result just obtained justifies our use of the interval $[\theta-n^{1/2+\varepsilon},\theta+n^{1/2+\varepsilon}]$ in place of $[\psi-n^{1/2+\varepsilon},\psi+n^{1/2+\varepsilon}]$. We will now expand the integral in the form $f(\theta+t)=e^{-c_2t^2}(c_0+c_1t+c_3t^3+\dots)$, where the c_i depend on $n, k, \varepsilon_4, \varepsilon_5, \dots$. $\phi(\theta)$

Consider the expansion $f(x)=e^{-x}x^n\exp(-\sum_{r=1}^{\infty}(w_r/rx^r))$. Since $w_r\leqslant n(4k)^r$ and $k=O(n^{1-\delta})$, we have that

$$\sum_{r=1}^{\infty}\frac{w_r}{rx^r}=\sum_{r=1}^{N}\frac{w_r}{rx^r}+O\left(\frac{k^7}{n^7}\right)$$

if $x\geqslant cn$ for some $c>0$. Now write $w_r=nu_r-\varepsilon_r=(nk^r/(r+1))\binom{2r}{r}+b_r$. By Lemmas 3.4 and 3.5, $b_r=O(nk^{r-1})$ for fixed r, and so uniformly for $r\leqslant N$. Therefore $\sum_{r=8}^{\infty}(b_r/rx^r)=O(k^7/n^7)$ if $x\geqslant cn$. Putting these together, we have that

$$\sum_{r=1}^{\infty}\frac{w_r}{rx^r}=n\sum_{r=1}^{N}\binom{2r}{r}\frac{k^r}{x^r(r+1)r}+\sum_{r=1}^{7}\frac{b_r}{rx^r}+O\left(\frac{k^7}{n^7}\right)\tag{1}$$

for $x\in[\theta-n^{1/2+\varepsilon},\theta+n^{1/2+\varepsilon}]$.

Next, write the integral J_k as $\phi(\theta)\int_{-n^{1/2+\varepsilon}}^{n^{1/2+\varepsilon}} f(0+t)/\phi(\theta)\,dt$. We begin with the factor

$$\phi(\theta)=e^{-\theta}\theta^n \exp\left(-n\sum_{r=1}^{\infty}\frac{u_r}{r\theta^r}\right)$$

$$=e^{-n}n^n \exp\left(n-\theta-n\log n/\theta-n\sum_{r=1}^{\infty}\frac{u_r}{r\theta^r}\right).$$

↓ 6·1(ii)

Using Lemmas 3.4(d), k the argument of the exponential can be expanded as a series with terms of the form ck^r/n^s, with $r\leqslant s+1$. The first few terms are

$$k-\frac{k^2-k}{2n}-\frac{k^3-3k^2+2k}{3n^2}-\cdots$$

Inspection suggests that the terms with $r=s+1$ are those of $n\log(1-k/n)$, and this can be proved by noting that $\theta=n^2/(n-k)+O(k/n)$ by Lemma 6.1(ii) and $u_r=\binom{2r}{r}k^r/(r+1)+O(k\quad)$ by Lemma 3.4(b). Thus $\phi(\theta)=e^{-n}n^n(1-k/n)^n\exp(\varphi_1(n,k))$, where $\varphi_1(n,k)$ has a convergent expansion in terms of the form ck^r/n^s with $1\leqslant r\leqslant s$.

The next problem is the integral $\int_{-n^{1/2+\varepsilon}}^{n^{1/2+\varepsilon}} f(0+t)/\phi(\theta)\,dt$. **A** simple manipulation gives that

$$\frac{f(\theta+t)}{\phi(\theta)}=\exp\left(-t+n\log(1+t/\theta)+\sum_{r=1}^{\infty}\frac{\varepsilon_r}{r(\theta+t)^r}\right.$$

$$\left.+n\sum_{r=1}^{\infty}\frac{u_r}{r\theta^r}\left(1-\left(1+\frac{t}{\theta}\right)^{-r}\right)\right)$$

$$=\exp\left(c_0+c_1t+c_2t^2+c_3t^3+c_4t^4+\cdots\right),$$

where $c_2=-1/2n-k(k-1)/2n^3+\ldots$, $c_3=1/3n^2+\ldots$, and similarly for the later terms. We can then expand $\exp(c_0+c_1t+c_3t^3+c_4t^4+\ldots)$ and integrate term-by-term using the well-known formula

$$\int_{-\infty}^{\infty}e^{-at^2}t^{2r}dt=\sqrt{\frac{\pi}{a}}\frac{(2r)!}{r!\,2^{2r}a^r}.$$

More than 200 terms need to be retained in the integrand to ensure the desired accuracy. Fortunately, the finiteness of Eq. (1) allows us no doubt as to which terms can be dropped.

The result of the integration is now multiplied by $\phi(\theta)$ to obtain the value of J_k, and the factor $e^{-n}n^n\sqrt{\pi/a}$ is converted to a radical-free form by means of Stirling's expansion. **At** last we obtain

6.2. THEOREM. *Let* $0 \leqslant k = 0(n^{1-\delta})$. *Then*

$$E_k = n!(1-k/n)^n \exp\left(\frac{k}{2n} + \frac{k(3k-1)}{6n^2} + \frac{k(2k^2-1)}{4n^3}\right.$$

$$+ \frac{k(15k^3+70k^2-105k+32)}{60n^4}$$

$$+ \frac{k(-6k^4+52k^3-68k^2+15k+9)}{12n^5}$$

$$+ \frac{k(-91k^5+441k^4-378k^3-406k^2+588k-148)}{42n^6}$$

$$\left. + \Delta + O\left(\frac{k^7}{n^7}\right)\right),$$

where

$$\Delta = \frac{\varepsilon_4}{4n^4} - \frac{(2k-3)\,\varepsilon_4}{2n^5} + \frac{\varepsilon_5}{5n^5}$$

$$+ \frac{(6k^2-36k+25)\,\varepsilon_4}{4n^6} - \frac{(k-2)\,\varepsilon_5}{n^6} + \frac{\varepsilon_6}{6n^6}$$

$$- \frac{(2k^3-44k^2+107k-45)\,\varepsilon_4}{2n^7} + \frac{(2k^2-14k+13)\,\varepsilon_5}{n^7}$$

$$- \frac{(2k-5)\,\varepsilon_6}{2n^7} + \frac{\varepsilon_7}{7n^7} + \frac{\varepsilon_4^2}{32n^8}.$$

Note that Theorem 6.2 implies that, asymptotically, a given $k \times n$ Latin rectangle R has $n!(1-k/n)^n$ extensions for $k = O(n^{1-\delta})$, irrespective of the structure of R. Stein proved this for $k = O(n^{2/3})$. (It follows from the van der Waerden bound that the number of extensions is at least $n!(1-k/n)^n$.)

As in Lemma 3.5, define s, h, and b to be the numbers of squares, hexagons, and $K_{2,3}$'s, respectively. With slightly less accuracy we can write Δ in terms of s, h, and b.

6.3 THEOREM. *If* $0 \leqslant k = O(n^{1-\delta})$ *then*

$$\Delta = \frac{s}{n^4} + \frac{2(2k-1)\,s}{n^5} + \frac{(10k^2-4k-11)\,s + h - 4b}{n^6} + O\left(\frac{k^6}{n^6}\right).$$

We can now appeal to the results of Section 4 to obtain an approximation for the average value $\bar{E}_k$ of the number of one-line extensions of a random $k \times n$ Latin rectangle.

6.4. THEOREM. *Let $0 \leqslant k = O(n^{1-\delta})$ for some 6 > 0. Then*

$$\bar{E}_k = n!(1-k/n)^n \exp\left(\frac{k}{2n} + \frac{k(3k-1)}{6n^2} + \frac{k(2k^2-1)}{4n^3} + \frac{k(30k^3+10k^2-15k-13)}{60n^4} + \frac{k(6k^4+4k^3-2k^2-9k+3)}{12n^5} + O\left(\frac{k^6}{n^6}\right)\right).$$

$$= n!\left(1 - \frac{k}{n} + \frac{k}{2n^2} - \frac{k}{6n^3} + \frac{k(7k-6)}{24n^4} + \frac{k(35k^2-10k-26)}{120n^5} + \frac{k(210k^3+75k^2-464k+180)}{720n^6} + O\left(\frac{k^6}{n^7}\right)\right)^n.$$

Proof. If a real random variable X is bounded above, then $\overline{\exp(X)} = \exp(\bar{X} + O(\overline{X^2}))$. Therefore, we can take the average within the exponential of Theorem 6.2 to the accuracy required.

The second expansion in Theorem 6.4 has the interesting feature that the coefficients of k^2/n^3, k^3/n^4, k^4/n^5, and k^5/n^6 are all zero. If this trend continues to similar later terms, the magnitude of the error term might actually be $O(k^5/n^7)$. In other words, the expansion may be considerably more accurate then we can prove.

From the equation $L(k, n) = \prod_{i=0}^{k-1} \bar{E}_i$ we obtain the following expansion.

6.5. THEOREM. *Let $0 \leqslant k = O(n^{1-\delta})$ for some 6 > 0. Then, uniformly,*

$$L(k, n) = (n!)^k \left(\frac{[n; k]}{n^k}\right)^n \exp(k(k-1)\, l(k, n)),$$

where

$$l(k, n) = \frac{1}{4n} + \frac{k-1}{6n^2} + \frac{k^2-k-1}{8n^3} + \frac{12k^3-13k^2-13k-6}{120n^4} + \frac{15k^4-18k^3-18k^2-28k+47}{180n^5} + O\left(\frac{k^5}{n^6}\right).$$

Theorem 6.5 yields the asymptotic value of $L(k, n)$ if $k = o(n^{6/7})$. As in Theorem 6.4, the leading coefficients of each term suggest, but do not prove, an estimate with considerably wider applicability.

The expansion in Theorem 6.5 can be converted to one of more familiar form:

$$L(k, n) = (n!)^k e^{-\binom{k}{2}} \times \left(1 - \frac{k(k-1)(k-2)}{6n} + \frac{k(k-1)^2(k-2)(k^2-2k-6)}{72n^2} + \ldots\right).$$

This is in agreement with Stein [27], who obtained the term $-k^3/6n$, but not with Erdős and Kaplansky [8]. The latter claim that the coefficient of $1/(72n^2)$ is $k(k-1)(k-2)(k^3-3k^2+8k-30)$, which is greater than our value by $12k(k-1)(k-2)(k-3)$. This minor error was also noted by Yamanoto [36]. It seems to occur in the derivation of Eq. (20) from Eq. (19) on page 235 of [8]. Also, if the method of Mendelsohn [22] is used to compute the number of one-row extensions of a $3\times 3n$ latin rectangle R with $G(R)$ consisting of n copies of $K_{3,3}$, then the result obtained agrees with ours.

For an application of Theorem 6.5 to random regular bipartite graphs, see [5].

7. Aftermath

We begin this section by mentioning a few other means we have used to verify our results. We believe this is important since parts of the calculation are too tedious to be verified by hand in any reasonable time. First, Theorem 6.2 was checked by numerically integrating $\phi(x)$ (see Lemma 6.1) for many values on n, k, $\varepsilon_4, ..., \varepsilon_7$. With the help of a numerical extrapolation technique [4], all the terms were verified, with accuracy ranging from 10^{-14} for the first term up to 10^{-1} for the least significant terms.

An excellent check of Theorem 6.5 is provided by comparing it, for $k=3$, with the asymptotic expansion derived by Yamamoto [32, 34]. This check should be sufficient to detect most single errors in the calculation, with the exception of the second term in Theorem 4.10 and the expression for ε_6 in Lemma 3.5. This lemma, together with Theorem 6.3, is consistent with the following result, of independent interest.

7.1. Theorem. *Let R be a $k\times n$ Latin rectangle such that $G(r)$ is n/k copies of $K_{k,k}$, where $k \mid n$ and $k=O(n^{1-\delta})$. Then the number of extensions of R is*

$$n!\left(1-\frac{k}{n}\right)^n \exp\left(\frac{k}{2n}+\frac{k(3k-1)}{6n^2}+\frac{k^2(3k-2)}{4n^3}\right.$$
$$+\frac{k(75k^3-80k^2+15k+2)}{60n^4}+\frac{k^2(26k^3-40k^2+11k+5)}{12n^5}$$
$$\left.+\frac{k(161k^5-336k^4+105k^3+98k^2-21k-1)}{42n^6}+O\left(\frac{k^7}{n^7}\right)\right).$$

Proof. The rook polynomial of $K_{k,k}$ is $L_k(x)$, the Laguerre polynomial degree of k normalized to be monic [14]. So $G(R)$ has rook polynomial $L_k(x)^{n/k}$ and the number of extensions of R is exactly

$$\int_0^{\infty} e^{-x} L_k(x)^{n/k}\, dx.$$

The entire calculation leading to Theorem 6.2 can now be repeated using the exact values of w_i $(i>0)$, yielding the stated result. As a check we note that $s=nk(k-1)^2/4$, $h=nk(k-1)^2\,(k-2)^2/6$, and $b=nk(k-1)^2\,(k-2)/6$. Substituting these values into Theorem 6.3, we obtain the same expansion within $O(k^6/n^6)$.

We next consider a much studied generalization of the "problème des ménages". Let $M(k, n)$ be the $k\times n$ Latin rectangle with rows $(1, 2, ..., n)$, $(2, 3, ..., n, 1), ..., (k, k+1, ..., n, 1, ..., k-1)$. Then the problem is to determine the number of extensions of $M(k, n)$. Exact solutions have been found for $k=2$ [28], $k=3$ [23, 24, 35], and $k=4$ [30]. Asymptotic expansions appear for $k=2$ in [26] (Problem 8.7(b)) and for $k\leqslant 3$ in [35]. Both expansions agree completely with the following result, and thus provide an additional check on our computations.

7.2. THEOREM. *If* $k=O(n^{1-\delta})$ *then the number of extensions of* $M(n, k)$ *is uniformly*

$$n!\left(1-\frac{k}{n}\right)^n \exp\left(\frac{k}{2n}+\frac{k(3k-1)}{6n^2}+\frac{k(8k^2-6k+1)}{12n^3}\right.$$
$$\left.+\frac{k(55k^3-70k^2+35k-8)}{60n^4}+O\left(\frac{k^5}{n^5}\right)\right).$$

Proof. If k is sufficiently large compared to n, then $G(M(k, n))$ has exactly $n\binom{k}{3}$ copies of C_4. The claim now follows from Theorem 6.3.

Finally, we give some conjectures arising from our work.

A. Theorems 6.2–6.5, 7.1, and 7.2 hold uniformly for $k\leqslant n/5$.

B. If $k=O(n^{1-\delta})$ then

$$L(k, n) \sim (n!)^k \left(\frac{[n; k]}{n^k}\right)^n \left(1-\frac{k}{n}\right)^{-n/2} e^{-k/2}.$$

We have proved this for $k=o(n^{6/7})$. It would be true as stated if the coefficient of k^{r-1}/n^r in $l(k, n)$ in Theorem 6.5 is $(2r+2)^{-1}$ for all r. Note that the conjecture is easily true in the logarithm. In fact,

$$\log L(k, n) \sim \log(n!)^k$$

uniformly for $0 \leqslant k \leqslant n$, since the van der Waerden lower bound and the naive upper bound $(n!)^k$ agree to that accuracy.

C. There are constants $0 \leqslant c_1 \leqslant c_2$ such that, for $0 \leqslant k \leqslant n$,

$$\exp(c_1 k^2/n) \leqslant L(k, n)(n!)^{-k} \left(\frac{[n; k]}{n^k}\right)^{-n} \leqslant \exp(c_2 k^2/n).$$

It is likely, but as yet unproven, that the lower bound can be derived from the inequality

$$E_k \geqslant I_k + \int_{4k}^{\infty} e^{-x}\phi(x)\, dx$$

in the notation of Sections 5 and 6.

D. For each t, n, the $t \times tn$ Latin rectangle with the greatest number of extensions is the one with $G(R)$ isomorphic to n copies of $K_{t,t}$. The evidence for this conjecture is Theorem 6.3, since the conjectured R uniquely maximises s.

REFERENCES

1. K. B. Athreya, C. R. Pranesachar, and N. M. Singhi, On the number of Latin rectangles and chromatic polynomial of $L(K_{r,s})$, *European J. Combin.* **1** (1980), 9–17.
2. R. Azor, J. Gillis, and J. D. Victor, Combinatorial applications of Hermite polynomials, *SIAM J. Math. Anal.* **13** (1982), 879–890.
3. N. Biggs, "Algebraic Graph Theory," Cambridge Univ. Press, London, 1974.
4. P. Bjørstad, G. Dahlquist, and E. Grosse, Extrapolation of asymptotic expansions by a modified Aitken δ^2-formula, *BIT* **21** (1981), 56–65.
5. B. Bollobás and B. D. McKay, The number of matchings in random regular graphs and bipartite graphs, *J. Combin. Theory, Ser. B* **41** (1986), 80–91.
6. P. J. Cameron, "Parallelisms of Complete Designs," London Mathematical Society Lecture Note Series, Vol. 23, Cambridge Univ. Press, London, 1976.
7. G. P. Egoritsjev, Solution of van der Waerden's permanent conjecture, preprint 13*M*, L. V. Kirensky Institute of Physics, 1980.
8. P. Erdős and I. Kaplansky, The asymptotic number of Latin rectangles, *Amer. J. Math.* **68** (1946), 230–236.
9. S. Even and J. Gillis, Derangements and Laguerre polynomials, *Math. Proc. Cambridge Philos. Soc.* **79** (1976), 135–143.
10. D. I. Falikman, A proof of the van der Waerden conjecture on the permanent of a doubly stochastic matrix, *Mat. Zametki* **29** (1981), 931–938.
11. I. Gessel, Counting Latin rectangles, *Bull. Amer. Math. Soc.* **16** (1987), 79–83.
12. C. D. Godsil, Matchings and walks in graphs, *J. Graphs Theory* **5** (1981), 285–297.
13. C. D. Godsil, Hermite polynomials and a duality relation for the matchings polynomial, *Combinatorica* **1** (1981), 257–262.
14. C. D. Godsil and I. Gutman, On the theory of the matching polynomial, *J. Graph Theory* **5** (1981), 137–144.

15. C. D. GODSIL AND B. D. MCKAY, Asymptotic enumeration of Latin rectangles, *Bull. Amer. Math. Soc.* **10** (1984), 91–92.
16. O. J. HEILMAN AND E. H. LIEB, Theory of monomer–dimer systems, *Comm. Math. Physics* **25** (1972), 190–232.
17. S. A. JONI AND G.-C. ROTA, A vector space analog of permutations with restricted position, *J. Combin. Theory, Ser. A* **29** (1980), 59–73.
18. B. D. MCKAY, The expected eigenvalue distribution of a large regular graph, *Linear Algebra Appl.* **40** (1981), 203–216.
19. B. D. MCKAY, Spanning trees in regular graphs, *European J. Combin.* **4** (1983), 149–160.
20. B. D. MCKAY, Asymptotics for 0–1 matrices with prescribed line sums, *in* "Enumeration and Design" (D. M. Jackson and S. A. Vanstone, Eds.), pp. 225–238, Academic Press, New York, 1984.
21. B. D. MCKAY AND N. C. WORMALD, The distribution of the number of cycles of moderate length in random regular graphs, to appear.
22. N. S. MENDELSOHN, The asymptotic series for a certain class of permutation problems, *Canad. J. Math.* **8** (1956), 234–245.
23. W. O. J. MOSER, The number of very reduced $4 \times n$ Latin rectangles, *Canad. J. Math.* **19** (1967), 1011–1017.
24. W. O. J. MOSER, A generalization of Riordan's formula for $3 \times n$ Latin rectangles, *Discrete Math.* **40** (1982), 311–313.
25. J. R. NECHVATAL, Asymptotic enumeration of generalized latin rectangles, *Utilitas Math.* **20** (1981), 273–292.
26. J. RIORDAN, "An Introduction to Combinatorial Analysis," Princeton Univ. Press, Princeton, NJ, 1978.
27. C. M. STEIN, Asymptotic evaluation of the number of Latin rectangles, *J. Combin. Theory, Ser. A* **25** (1978), 38–49.
28. J. TOUCHARD, Sur un problème de permutations, *C. R. Acad. Sci. Paris* **198** (1934), 631–633.
29. J. H. VAN LINT, Notes on Egoritsjev's proof of the van der Waerden conjecture, *Linear Algebra Appl.* **39** (1981), 1–8.
30. E. G. WHITEHEAD, Four-discordant permutations, *J. Austral. Math. Soc. Ser. A* **28** (1979), 369–377.
31. N. C. WORMALD, "Some Problems in the Enumeration of Labelled Graphs," Ph.D. Thesis, Department of Mathematics, University of Newcastle, 1978.
32. K. YAMAMOTO, An asymptotic series for the number of three-line Latin rectangles, *J. Math. Soc. Japan* **2** (1949), 226–241.
33. K. YAMAMOTO, On the asymptotic number of Latin rectangles, *Japan. J. Math.* **21** (1951), 113–119.
34. K. YAMAMOTO, Symbolic methods in the problem of three-line Latin rectangles, *J. Math. Soc. Japan* **5** (1953), 13–23.
35. K. YAMAMOTO, Structure polynomial of Latin rectangles and its application to a combinatorial problem, *Mem. Fac. Sci. Kyushu Univ. Ser. A* **10** (1956), 1–13.
36. K. YAMAMOTO, On the number of Latin rectangles, *Res. Rep. Sci. Div. Tokyo Womens' Univ.* **19** (1969), 86–97.

Printed by Catherine Press, Ltd., Tempelhof 41, B-8000 Brugge, Belgium