

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 $\{1, 2, \dots, n\}$ 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 $n \rightarrow \infty$ 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 $\{1, \dots, n\}$ 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.

determine the asymptotic value of $L(k, n)$ as $n \rightarrow \infty$, 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 $k = O((\log n)^{3/2-\epsilon})$,

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

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

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

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

In this paper we prove

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

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

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

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

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 $k \leq 3$ 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 $\{c_1, \dots, c_n\}$ and $\{e_1, \dots, e_n\}$. A vertex c_i is adjacent to a vertex e_j iff the integer j occurs in the i th column of L . Note that, for each m ($1 \leq m \leq k$), the edges $\{c_i, e_j\}$ such that j is in the m th row of the i th 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 $r = 1, 2, \dots$. 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, \dots, 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 $\mathbf{v} = (v_0, \dots, v_r)$ is *tree-like* if, for each $i = 0, \dots, r$, the walk (v_0, \dots, 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 \mathcal{M} , 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 $\lambda_1, \dots, \lambda_n$ 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 $x \geq 4k - 4$ 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 ,

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

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

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

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 $n \rightarrow \infty$ 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 $m \leq 3$, so fix $m \geq 4$. 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 $1 \leq i \leq k$ and $1 \leq j, \alpha \leq n$. 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 $\text{pos}(x)$ and the latter by $\text{cont}(x)$. 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 $L \setminus H$. Then we have:*

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

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

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 $r(L) \leq n - 4k$ 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 $L \setminus H$ 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 $r_i(H) \leq n - 1$ we have $N(L, H \cup \{x\})(n - 2k + 1 - r_i(H)) \leq M \leq N(L, H \cup \{x\})(n - r_i(H) - 1)$.*

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 $R_2 \setminus (H \cup \{x\})$ 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*

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

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 $\text{pos}(x_1) = \text{pos}(x)$. Let y be the unique triple in R_1 in the same row as x (and x_1) with $\text{cont}(y) = \text{cont}(x)$. 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

$M \leq N(L, H)$. 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, $M \leq N(L, H)$ it follows, using the lower bound in Lemma 4.2, that

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

This holds for all L, H satisfying the given conditions and so can also be applied with $H \cup \{x\}$ 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 $|A| \leq N(L, H)(n-2k-r(H))^{-1}$.

We thus have a bound on the probability that x_2 contains $\text{cont}(x)$. The probability that some triple in the x -column contains $\text{cont}(x)$ 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 $r(L) \leq n-4k$, 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

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

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

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

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 $v > 4$, $r(L) \leq n-vk$. Then the probability $P(k, n, L)$ that a random $k \times n$ Latin rectangle contains L is*

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

