THE ASYMPTOTIC NUMBER OF LATIN RECTANGLES.*

By Paul Erdös and Irving Kaplansky

1. Introduction. The problem of enumerating n by k Latin rectangles was solved formally by MacMahon [4] using his operational methods. For $k=3$, more explicit solutions have been given in [1], [2], [3], and [5]. While further exact enumeration seems difficult, it is an easy heuristic conjecture that the number of n by k Latin rectangles is asymptotic to $(n!)^k \exp(-{}_kC_2)$. Because of an error, Jacob [2] was led to deny this conjecture for $k=3$; but Kerawala [3] rectified the error and then verified the conjecture to a high degree of approximation. The first proof for $k=3$ appears to have been given by Riordan [5].

In this paper we shall prove the conjecture not only for k fixed (as $n \to \infty$) but for $k < (\log n)^{3/2-\epsilon}$. As indicated below, a considerably shorter proof could be given for the former case. The additional detail is perhaps justified by (1) the interest attached to an approach to Latin squares ($k=n$), (2) the emergence of further terms of an asymptotic series (**4**), (3) the fact that $(\log n)^{3/2}$ appears to be a "natural boundary" of the method. (We believe however that the actual break occurs at $k=n^{1/3}$.)

2. Notation. An n by k Latin rectangle L is an array of n rows and k columns, with the integers $1, \cdots, n$ in each row and all distinct integers in each column. Let N be the number of ways of adding a $(k+1)$-st row to L so as to make the augmented array a Latin rectangle. We use the sieve method (method of inclusion and exclusion) to obtain an expression for N. From $n!$, the total number of possible choices for the $(k+1)$-st row, we take away those having a clash with L in a given column—summed over all choices of that column, then reinstate those having clashes in two given columns, etc. The result can be written

$$N = \sum_{r=0}^{n} (-)^r A_r (n-r)! \tag{1}$$

where A_r is the number of ways of choosing r distinct integers in L, no two in the same column. In particular $A_0 = 1$, $A_1 = nk$. To estimate the higher values of A_r we apply the sieve method again. The total number of ways of

* Received November 30, 1945.

selecting r elements of L, not necessarily distinct integers but with no two in the same column, is ${}_nC_r k^r$. This over-estimates A_r; we have to take away those selections which include a specified pair of 1's, 2's, $\cdots$, or n's, then reinstate those which include two pairs, etc. We may write the result

$$A_r = \sum_s (-)^s B(r,s). \tag{2}$$

Here $B(r,s)$ is precisely defined as follows. Take any s of the $n\,{}_kC_2$ pairs of 1's, $\cdots$, n's which can be formed in L. Suppose that this selection involves in all y elements; y may be as large as $2s$, or as small as the integer for which ${}_yC_2 = s$. Find the number of ways of adjoining $r - y$ further elements, so as to form a set of r elements with no two in the same column. The result of summing over all choices of s pairs is, by definition, $B(r,s)$. We note in particular that

$$B(r,0) = {}_nC_r\, k^r \tag{3}$$

$$B(r,1) = n\,{}_kC_2\, {}_{n-2}C_{r-2}\, k^{r-2}. \tag{4}$$

The B's may be analyzed further as follows. Let $F(s,t)$ be the number of ways of choosing s pairs of 1's, $\cdots$, n's, which use up t elements in all, and for which no two of the t elements lie in the same column. The number of ways of expanding this selection of t elements to r elements, with no two in the same column, is ${}_{n-t}C_{r-t}\, k^{r-t}$. Hence

$$B(r,s) = \sum_t F(s,t)\, {}_{n-t}C_{r-t}\, k^{r-t}. \tag{5}$$

It is to be observed that extreme limits for the summation in (5) are given by $t \leqq 2s$ and $s \leqq {}_tC_2$ or, more generously, $\sqrt{s} \leqq t$.

These quantities $F(s,t)$ are the ultimate building blocks from which the exact value of N is constructed. We shall discuss them further in **4**. For the present the following crude inequality will suffice:

$$\sum_s F(s,t) < n^{t/2}(k^2 t)^{t^2}. \tag{6}$$

The proof of (6) is as follows. The left hand side is just the number of ways of choosing a set of (any number of) pairs which involve in all precisely t elements. In such a choice at most $[t/2]$ distinct integers are permissible, and these may be taken in less than $n^{t/2}$ ways. In all we have at most

${}_tC_2 < t^2$ pairs to dispose of in the selection. For each of these t^2 pairs we have ${}_kC_2\, t/2 < k^2t$ possibilities and hence for all of them at most $(k^2t)^{t^2}$ choices. This establishes (6).

The various quantities defined in this section will be used without further explanation in the remainder of the paper.

3. Proof of the main result. We first prove

THEOREM 1. *If $k < \log n)^{3/2-\epsilon}$, then for sufficiently large n*

$$|Ne^k/n! - 1| < n^{-c} \tag{7}$$

where c is a positive constant depending only on ϵ.

Proof. Define $A(r,x)$ by

$$A(r,x) = \sum_{s=1}^{x-1} (-)^s B(r,s), \tag{8}$$

where $x = [(\log n)^{1-\epsilon}]$. Then by the sieve's well known property of being alternately in excess and defect we have

$$|A_r - B(r,0) - A(r,x)| \leqq B(r,x). \tag{9}$$

In (1) make the substitution

$$A_r = \{A_r - B(r,0) - A(r,x)\} + B(r,0) + A(r,x)$$

and use (3) and (9). We find

$$|N - \sum_{r=0}^{n} (-)^r {}_nC_r k^r (n-r)!| \leqq |G| + H, \tag{10}$$

where

$$G = \sum_{r=0}^{n} (-)^r A(r,x)(n-r)!, \tag{11}$$

$$H = \sum_{r=0}^{n} B(r,x)(n-r)!. \tag{12}$$

We proceed to study G. With the use of (8) and (5), and an interchange of summation signs, (11) becomes

$$G/n! = \sum_{s=1}^{x-1} (-)^s \sum_t F(s,t) \sum_{r=t}^{n} (-)^r \, {}_{n-t}C_{r-t} \, k^{r-t}/(n)_r$$

where $(n)_r = n(n-1)\cdots(n-r+1)$ is the Jordan factorial notation. The change of variable $r = t + u$ transforms the final sum into

$$(-)^t/(n)_t \sum_{u=0}^{n-t} (-k)^u/u! = (-)^t(e^{-k} - \theta)/(n)_t$$

where θ is the remainder after $n - t$ terms of the series for e^{-k}. Then

$$|G| \, e^k/n! \leqq \sum_{s=1}^{x-1} \sum_t F(s,t)(1 + \theta e^k)/(n)_t. \tag{13}$$

As noted above, the limits for t lie between $\sqrt{s}$ and $2s$. Hence $t \leqq 2x < 2 \log n$. From this we readily deduce

$$1/(n)_t < c_1 n^{-t}, \tag{14}$$

$$\theta e^k < c_2, \tag{15}$$

where c_1, c_2 are absolute constants. From (6), (13), (14), and (15) we obtain

$$|G| \, e^k/n! < c_3 \sum_{t=1}^{2x} (k^2 t)^{t^2}/n^{t/2}$$

with $c_3 = c_1(1 + c_2)$. In the fraction under the summation sign, the logarithms of numerator and denominator are respectively of the orders $t^2 \log \log n$ and $t \log n$. Since $t < 2(\log n)^{1-\epsilon}$, it follows that for large n

$$(k^2 t)^{t^2}/n^{t/2} < n^{-c_4}$$

where c_4 is a positive constant depending only on ϵ. Hence

$$|G| \, e^k/n! < 2x c_3 n^{-c_4} < n^{-c_5}. \tag{16}$$

We next turn our attention to the term H given by (12). From (5) and an interchange of orders of summation,

$$H/n! = \sum_t F(x,t) \sum_{r=t}^{n} {}_{n-t}C_{r-t} \, k^{r-t}/(n)_r.$$

The final sum is the product of $1/(n)_t$ by a portion of the series for e^k. Hence

$$H/n! < e^k \sum_t F(x,t)/(n)_t < c_1 e^k \sum_t (k^2 t)^{t^2}/n^{t/2}$$

by (6) and (14). The fraction to be estimated is the same as above but the summation now starts at $\sqrt{x} \geqq c_6(\log n)^{(1-\epsilon)/2}$. It follows that $t \log n \geqq c_8(\log n)^{3/2-\epsilon/2}$, and we are able to swallow up a further term e^{2k} whose logarithm is less than $2(\log n)^{3/2-\epsilon}$. Hence for large n

$$e^{2k}(k^2 t)^{t^2}/n^{t/2} < n^{-c_7}$$

and

$$(17) \qquad He^k/n! < 2xc_1 n^{-c_7} < n^{-c_9}.$$

Combining (16), (17), and (10), we obtain (7), for the sum on the left of (10) may run to infinity at a cost of $O(n^{-c})$. This concludes the proof.

(We may note that for the case where k is fixed as $n \to \infty$, the proof could be abridged as follows. We take $x = 1$; then the term G disappears, and an estimate of H is easily obtained from (4).)

From Theorem 1 we readily derive our main result:

THEOREM 2. *Let $f(n,k)$ be the number of n by k Latin rectangles and suppose $k < (\log n)^{3/2-\epsilon}$. Then*

$$(18) \qquad f(n,k)(n!)^{-k} \exp({}_kC_2) \to 1 \quad \textit{as} \quad n \to \infty.$$

Proof. From Theorem 1 it follows that $f(n, i+1)$ lies between the limits $f(n,i)n!\,e^{-i}(1 \pm n^{-c})$. Taking the product from $i = 1$ to $k - 1$, we find that $f(n,k)$ lies between the limits

$$(n!)^k \exp(-{}_kC_2)(1 \pm n^{-c})^k.$$

Since $(1 + n^{-c})^k$ and $(1 - n^{-c})^k \to 1$ as $n \to \infty$, we obtain (18).

4. Further terms of the asymptotic series. A more careful argument reveals that the error term in (7) is actually of the order of $k^2 n^{-1}$. By detaching the term $B(r,1)$ as well as $B(r,0)$ in (2), we can reduce the error to the order of $k^4 n^{-2}$. Continuing in this fashion, we may compute successive terms of an asymptotic series. The existence of such a series was conjectured by Jacob [2, 337].

We shall merely sketch the results. Applying (1), (2), and (5) as we did in **3**, we find

$$N/n! = \sum_s (-)^s \sum_t F(s,t)(e^{-k} - \theta)/(n)_t.$$

The term θ may be dropped and we have

$$(19)\qquad Ne^k/n! = 1 - \frac{F(1,2)}{(n)_2} + \frac{F(2,3)}{(n)_3} + \frac{F(2,4)}{(n)_4} - \cdots$$

Thus all that is required is evaluation of the F's. That $F(1,2) = n\,{}_kC_2$ was already implicitly noted in (4). For $F(2,3)$ we observe that not more than one integer may be used, that there are then $n\,{}_kC_3$ choices for the three elements, and 3 choices for the two pairs within them. Hence $F(2,3) = 3n\,{}_kC_3$. Similarly $F(2,4)$ includes the term $3n\,{}_kC_4$, corresponding to the choice of only one integer. If two different integers are taken, there are *ab initio* ${}_nC_2({}_kC_2)^2$ choices; but we must eliminate selections which include two elements in the same column. An application of the sieve process to this last difficulty yields

$$F(2,4) = 3n\,{}_kC_4 + {}_nC_2({}_kC_2)^2 - n\,{}_kC_2(k-1)^2 + X,$$

where X is the number of instances in which integers i, j both occur in two different columns. It is noteworthy that this is the first term which depends upon the particular Latin rectangle to which a $(k+1)$-st row is being added.

A simple argument shows that $X \leqq n\,{}_kC_2(k-1)$, so that $X/(n)_4$ is of order n^{-3} or less, as are all the later terms of (19). Hence we have, correct up to n^{-2}:

$$(20)\qquad \begin{aligned} Ne^k/n! &= 1 - \frac{n\,{}_kC_2}{(n)_2} + \frac{2n\,{}_kC_3}{(n)_3} + \frac{{}_nC_2({}_kC_2)^2}{(n)_4} + \cdots \\ &= 1 - {}_kC_2/n + {}_kC_2(k+4)(3k-7)/12n^2 + \cdots. \end{aligned}$$

By taking the product of the terms (20) from 1 to $k-1$, we obtain the asymptotic series for $f(n,k)$, the number of Latin rectangles:

$$(21)\qquad \begin{aligned} &f(n,k)\,(n!)^{-k}\exp({}_kC_2) \\ &\quad = 1 - {}_kC_3/n + {}_kC_3(k^3 - 3k^2 + 8k - 30)/12n^2 + \cdots. \end{aligned}$$

For $k = 3$, the right side of (21) becomes $1 - 1/n - 1/2n^2 + \cdots$. In the table below we compare this with the exact value given by Kerawala in [3].

n	$1 - 1/n - 1/2n^2$	Exact value of (21)
5	.78	.76995
10	.895	.89560
15	.93111	.93126
20	.94875	.94881
25	.9592	.95923

In attempting to push the asymptotic series still further, we run into the difficulty that terms like X, i. e., terms dependent upon the preceding Latin rectangle, begin to play a rôle in (20). However, it may be that in (21) at least the term in n^{-3} can be obtained without consideration of X, for heuristically it seems likely that the "expectation" of X is $o(n)$.

In conclusion we remark that the form of (21) strongly suggests that at about $k = n^{1/3}$ the expression ceases to be valid. We are unable to prove this rigorously.

STANFORD UNIVERSITY
AND
UNIVERSITY OF CHICAGO.

BIBLIOGRAPHY

1. L. Dulmage, Problem E 650, *American Mathematical Monthly*, vol. 51 (1944), pp. 586-587.

2. S. M. Jacob, "The enumeration of the Latin rectangle of depth three . . .", *Proceedings of the London Mathematical Society*, vol. 31 (1930), pp. 329-354.

3. S. M. Kerawala, "The enumeration of the Latin rectangle of depth three by means of a difference equation," *Bulletin of the Calcutta Mathematical Society*, vol. 33 (1941), pp. 119-127.

4. P. A. MacMahon, *Combinatory Analysis*, vol. I, Cambridge, 1915, pp. 246-263.

5. J. Riordan, "Three-line Latin rectangles," *American Mathematical Monthly*, vol. 51 (1944), pp. 450-452.