Subsquares in random Latin rectangles

Jack Allsop Ian M. Wanless

School of Mathematics
Monash University
Vic 3800, Australia

jack.allsop@monash.edu ian.wanless@monash.edu

Abstract

Suppose that k is a function of n and $n \to \infty$. We show that with probability $1-O(1/n)$, a uniformly random $k \times n$ Latin rectangle contains no proper Latin subsquare of order 4 or more, proving a conjecture of Divoux, Kelly, Kennedy and Sidhu. We also show that the expected number of subsquares of order 3 is bounded and find that the expected number of subsquares of order 2 is $\binom{k}{2}(1/2 + o(1))$ for all $k \leqslant n$.

1 Introduction

Throughout this paper, n and k will denote positive integers with $k \leqslant n$ and all asymptotics are as $n \to \infty$, with k some function of n. All probability distributions will be discrete and uniform, with $\Pr(\cdot)$ denoting probability.

A $k \times n$ *Latin rectangle* L is a $k \times n$ matrix on n symbols, each of which occurs at most once in each row and column. If $k = n$ then L is a *Latin square* of order n. A *subsquare* of L is a submatrix of L which is itself a Latin square. Every $k \times n$ Latin rectangle has kn subsquares of order 1, and every Latin square of order n has a single subsquare of order n. A *proper* subsquare of L is a subsquare which has order in the set $\{2, 3, \dots, n-1\}$. The order of any proper subsquare of L is at most $\min\{k, n/2\}$. An *intercalate* is a subsquare of order 2.

Let m be an integer satisfying $2 \leqslant m \leqslant n/2$ and define $\mathbb{E}_m(k, n)$ to be the expected number of subsquares of order m in a random $k \times n$ Latin rectangle. The main result of this paper is the following theorem.

Theorem 1.1. *If m, k are integer functions of n satisfying $4 \leqslant m \leqslant \min\{k, n/2\}$ then $\mathbb{E}_m(k, n) = O(n^{-2})$ as $n \to \infty$.*

We also show that $\mathbb{E}_3(k, n) = O(k/n)$. An immediate corollary of Theorem 1.1 is the following result, proving a conjecture of Divoux, Kelly, Kennedy and Sidhu [3], which generalises a conjecture by McKay and Wanless [9].

Corollary 1.2. *With probability $1-O(1/n)$, a random $k \times n$ Latin rectangle has no proper subsquare of order 4 or more.*

The structure of this paper is as follows. In §2 we discuss some of the literature regarding subsquares in random Latin squares and rectangles, as well as giving some definitions which we require to prove Theorem 1.1. We prove Theorem 1.1 in §3, and in §4 we show that $\mathbb{E}_3(k, n) = O(k/n)$ and $\mathbb{E}_2(k, n) = (1/2 + o(1))\binom{k}{2}$ for all $k \leqslant n$.

2 Background

A seminal paper by Godsil and McKay [5] used switching to count substructures in random Latin rectangles. A special case of their Theorem 4.7 is that if $k \leqslant n/6$ then the probability that a random $k \times n$ Latin rectangle contains a specific subsquare of order m is $\exp(-m^2 \log n + O(m^2 k/n))$. There are less than $\binom{k}{m}\binom{n}{m}^2 \leqslant (en/m)^{3m}$ ways to choose the rows, columns and symbols for a potential subsquare of order m and less than m^{m^2} Latin squares with those rows, columns and symbols. It follows that if $k \leqslant n/6$ then

$$\mathbb{E}_m(k,n) \leqslant \exp((m^2 - 3m)\log(m/n) + O(m) + O(m^2 k/n))$$

and, using precise values rather than approximations,

$$\mathbb{E}_2(k,n) = \frac{1}{2}\binom{k}{2}\exp(O(k/n)). \tag{2.1}$$

In particular, if $k = o(n)$ then a random $k \times n$ Latin rectangle almost surely has no subsquare of order 4 or larger.

The first researchers to study subsquares of random Latin squares were McKay and Wanless [9]. They provided some estimates for the number of intercalates and conjectured that asymptotically almost surely there will be no proper subsquare of order 4 or more. They also conjectured that $\mathbb{E}_3(n,n) = 1/18 + o(1)$. They proved that with probability $1 - o(1)$, a random Latin square of order n has no subsquare of order $n/2$. Estimates of the number of intercalates were subsequently improved in [2, 6, 7, 8], leading to the following result.

Theorem 2.1. *Let* $\mathbf{N}$ *denote the number of intercalates in a random Latin square of order* n*.*

- *The expected value of* $\mathbf{N}$ *is* $(1 + o(1))n^2/4$*,*
- $\Pr(\mathbf{N} \leqslant (1-\delta)n^2/4) = \exp(-\Theta(n^2))$ *for every* $\delta \in (0, 1]$*,*
- $\Pr(\mathbf{N} \geqslant (1+\delta)n^2/4) = \exp(-\Theta(n^{4/3}\log n))$ *for every* $\delta > 0$*.*

After two decades of no progress on the McKay-Wanless conjectures, two research groups recently managed to settle the case of large subsquares. Divoux, Kelly, Kennedy and Sidhu [3] proved that there is a constant $K > 0$ such that asymptotically almost all Latin squares of order n contain no proper subsquare of order $K(n \log n)^{1/2}$ or more. They also showed that for $k \leqslant (1/2 - o(1))n$, a random $k \times n$ Latin rectangle contains no proper subsquare of order 4 or more with probability $1 - o(1)$, and that $\mathbb{E}_3(k,n) = (1 + o(1))\binom{k}{3}/(3n^3)$ and $\mathbb{E}_2(k,n) = \binom{k}{2}(1/2 + o(1))$. They conjectured that these results hold for all $k \leqslant n$, generalising the conjectures made by McKay and Wanless [9].

Independently, Gill, Mammoliti and Wanless [4] showed that for any $\varepsilon > 0$, asymptotically almost all Latin squares contain no proper subsquare of order $n^{1/2}\log^{1/2+\varepsilon} n$ or more. They also proved that their result implies that isomorphism for Latin squares can be tested in average case polynomial time.

We now give some definitions which we will need to prove Theorem 1.1. Let u and v be positive integers. We denote the set $\{1, 2, \ldots, u\}$ by $[u]$. Let A be a $u \times v$ matrix. We will index the rows of A by $[u]$ and the columns of A by $[v]$. This convention will be adopted throughout this paper, unless stated otherwise. A pair $(i,j) \in [u] \times [v]$ is a *cell* of A and a triple $(i, j, A_{i,j})$ is an *entry* of

$$\begin{bmatrix} 2 & 3 & 5 & 6 & 1 & 4 & 7 \\ 5 & 4 & 7 & 2 & 3 & 6 & 1 \\ 4 & 5 & 1 & 3 & 7 & 2 & 6 \\ 6 & 1 & 3 & 4 & 2 & 7 & 5 \end{bmatrix}$$

Figure 1: An incomplete column cycle in a 4×7 Latin rectangle

A. Let $I \subseteq [u]$ and $J \subseteq [v]$. We denote by $A[I, J]$ the submatrix of A induced by the rows in I and the columns in J.

A $k \times n$ *partial Latin rectangle* P is a $k \times n$ matrix such that each cell is either empty or contains one of n symbols, such that no symbol occurs more than once in each row and column. A *subrectangle* of P is a submatrix of P which is itself a Latin rectangle. We will equivalently view P as the set of triples (r, c, s) where (r, c) is a non-empty cell of P and $s = P_{r,c}$. This allows us to use set notation such as $(1, 2, 3) \in P$, which means that $P_{1,2} = 3$. Throughout this paper, unless otherwise stated, a $k \times n$ partial Latin rectangle will have symbol set $[n]$. The convention that a Latin square of order n has row indices, column indices and symbol set $[n]$ will be broken when dealing with subsquares.

We will prove Theorem 1.1 using the switching method. This method is behind many of the results on random Latin squares and rectangles including [2, 3, 4, 5, 6, 7, 8, 9] and has been widely used for many other combinatorial structures. Our switching procedure will be cycle switching and incomplete cycle switching, which we define now.

Let P be a $k \times n$ partial Latin rectangle. A *row cycle* of P is a $2 \times \ell$ subrectangle, for some $\ell \leqslant n$, which is minimal in the sense that it does not contain any $2 \times \ell'$ subrectangle with $\ell' < \ell$. Let Q be a row cycle of P which hits rows i, j and the columns in C, for some $\{i, j\} \subseteq [k]$ and $C \subseteq [n]$. Since Q is uniquely determined by rows i, j and a single column in C, we denote Q by $\rho(i, j, c)$, where c is any element of C. Row cycles give us a way of perturbing partial Latin rectangles to create new ones. We can define a partial Latin rectangle P' by

$$P'_{x,y} = \begin{cases} P_{i,y} & \text{if } x = j \text{ and } y \in C, \\ P_{j,y} & \text{if } x = i \text{ and } y \in C, \\ P_{x,y} & \text{otherwise.} \end{cases}$$

We say that P' has been obtained from P by switching on $\rho(i, j, c)$.

A *column cycle of length* ℓ is an $\ell \times 2$ submatrix of P that contains exactly ℓ symbols and no empty cells and does not contain any smaller submatrix with these properties. The entries in a column cycle form a row cycle when P is transposed. Suppose that $\mathcal{C}$ is an $\ell \times 2$ submatrix of P that does not contain a column cycle or any empty cells and which hits columns i and j and rows R of P. Then $\mathcal{C}$ is an *incomplete column cycle of length* ℓ if there are unique rows $r, r' \in R$ such that $P_{r,i}$ does not occur in column j of P and $P_{r',j}$ does not appear in column i of P. See Figure 1 for an example of an incomplete column cycle.

Let $\{i, j\} \subseteq [n]$ with $i \neq j$ and let $r \in [k]$. There is either a unique column cycle or a unique incomplete column cycle (but not both) which hits columns i and j and row r. Denote this column cycle or incomplete column cycle by $\sigma(i, j, r)$, and let R be the set of rows which it hits. The substructure $\sigma(i, j, r)$ gives us a way of perturbing P to create a new Latin rectangle. Define a partial Latin rectangle P' by

$$P'_{x,y} = \begin{cases} P_{x,i} & \text{if } x \in R \text{ and } y = j, \\ P_{x,j} & \text{if } x \in R \text{ and } y = i, \\ P_{x,y} & \text{otherwise.} \end{cases}$$

We say that P' has been obtained from P by switching on $\sigma(i, j, r)$.

There are also symbol cycles and incomplete symbol cycles, which are, respectively, the image of column cycles and incomplete column cycles under the map which replaces each entry (r, c, s) by the entry (r, s, c). Symbol cycles and incomplete symbol cycles can also be switched in an analogous way. See [10] for a study of switching on row, column and symbol cycles.

Let L be a $k \times n$ Latin rectangle. Let $\{i, j\} \subseteq [k]$ with $i \neq j$. The permutation mapping row i of L to row j, denoted by $\tau_{i,j}$, is defined by $\tau_{i,j}(L_{i,\ell}) = L_{j,\ell}$ for every $\ell \in [n]$. Such permutations are called *row permutations* of L. Let ρ be a cycle in $\tau_{i,j}$ and in row i let C be the set of columns containing the symbols involved in ρ. Then the set of entries

$$\{(i, c, L_{i,c}), (j, c, L_{j,c}) : c \in C\}$$

is the row cycle $\rho(i, j, c)$ of L, where c is any element of C. Conversely, every row cycle of L corresponds to a cycle of a row permutation of L.

Now suppose that $\{i, j\} \subseteq [n]$ with $i \neq j$. The partial permutation mapping column i to column j, denoted by $\kappa_{i,j}$, is defined by $\kappa_{i,j}(L_{\ell,i}) = L_{\ell,j}$ for every $\ell \in [k]$. A cycle in $\kappa_{i,j}$ corresponds to a column cycle of L hitting columns i and j. Say that a list $[x_1, x_2, \ldots, x_u]$ is an *incomplete cycle* of $\kappa_{i,j}$ if $\kappa_{i,j}(x_i) = x_{i+1}$ for every $i \in [u-1]$ and $\kappa_{i,j}(x_u)$ and $\kappa_{i,j}^{-1}(x_1)$ are undefined. An incomplete cycle of $\kappa_{i,j}$ corresponds to an incomplete column cycle of L hitting columns i and j.

3 Proof of main theorem

Throughout this section $m = m(n)$ will be a positive integer satisfying $2 \leqslant m \leqslant \min\{k, n/2\}$. This section is split into three subsections. In §3.1 and §3.2 we prove two different bounds on the probability that a random $k \times n$ Latin rectangle contains a subsquare of order m in a specific selection of rows and columns, and on a specific symbol set. These bounds are effective for different but overlapping ranges of m. In §3.3 we combine them to prove Theorem 1.1, using the bound from §3.1 for $m \geqslant 6$ and the bound from §3.2 for $m \in \{4, 5\}$.

3.1 Bounding probability of large subsquares

The main result of this subsection is the following theorem, which gives a nontrivial bound for all m. However, for $m = 4$ it is not strong enough to imply any result in the direction of Theorem 1.1 and for $m = 5$ it is not enough to imply the full strength of Theorem 1.1. The bound we prove in the next subsection will be better whenever $2 < m = O(1)$.

Theorem 3.1. *Let $R \subseteq [k]$ be of cardinality m, let C and S be subsets of $[n]$ of cardinality m, and let $\mathbf{L}$ be a random $k \times n$ Latin rectangle. The probability that $\mathbf{L}[R, C]$ is a Latin square of order m with symbol set S is at most*

$$\frac{kn(k+1-m)(n+1-m)}{m^2 \binom{n}{m}^3 \binom{k}{m}}.$$

Without loss of generality we may assume that $R = C = S = [m]$ in Theorem 3.1. Let $T_0 = \varnothing$ and for $i \in [m^2]$ define T_i to be the set of all pairs $(r, c) \in R \times C$ such that $(c-1)m + r \leqslant i$. We will consider the sets T_i to be sets of cells of $k \times n$ Latin rectangles. So if $a \in \{0, 1, \ldots, m-1\}$, $b \in [m]$ and L is a $k \times n$ Latin rectangle then T_{am+b} consists of all cells of L in the first m rows and a columns, or in the first b rows and the $(a+1)$-st column. Define t_i to be the cell in $T_i \setminus T_{i-1}$. Define Δ_i to be the set of $k \times n$ Latin rectangles such that the symbol in cell (r, c) is an element of $[m]$, for every $(r, c) \in T_i$. Note that Δ_0 is the set of all $k \times n$ Latin rectangles.

Let $\mathbf{L}$ be a random $k \times n$ Latin rectangle. The probability that $\mathbf{L}[[m], [m]]$ is a Latin square of order m with symbol set $[m]$ is $\Pr(\mathbf{L} \in \Delta_{m^2})$. By the chain rule of probability,

$$\Pr(\mathbf{L} \in \Delta_{m^2}) = \prod_{i=1}^{m^2} \Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}). \tag{3.1}$$

Similar to the approach of Divoux, Kelly, Kennedy and Sidhu [3], our approach to proving Theorem 3.1 is to bound the terms

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}). \tag{3.2}$$

However we will only provide a non-trivial bound on (3.2) when the cell $t_i = (r, c)$ satisfies $\{1, m\} \cap \{r, c\} \neq \varnothing$. That is, we only bound (3.2) when t_i lies in the first row, first column, m-th row or m-th column. Slightly surprisingly, this turns out to be enough to derive our result. Consider when $k = 4$, $n = 6$, $m = 3$ and $i = 5$. The Latin rectangle

$$L = \begin{bmatrix} 1 & 2 & 6 & 5 & 4 & 3 \\ 2 & 3 & 5 & 4 & 6 & 1 \\ 3 & 4 & 1 & 2 & 5 & 6 \\ 4 & 6 & 2 & 1 & 3 & 5 \end{bmatrix}$$

is a member of Δ_i, due to the highlighted entries. Suppose that we want to use column cycle or incomplete column cycle switching to get from L to a rectangle in $\Delta_{i-1} \setminus \Delta_i$. Clearly we must switch on $\sigma(2, c, 2)$ for some $c \in \{3, 4, 5, 6\}$. If we take $c = 5$ it is clear that the resulting rectangle will lie in $\Delta_{i-1} \setminus \Delta_i$, as desired. However, if we take $c = 4$ the resulting rectangle will not be a member of Δ_{i-1}, since cell $(1, 2)$ will no longer contain a symbol in $[3]$. So to estimate the number of switches from some rectangle $L' \in \Delta_i$ to a rectangle in $\Delta_{i-1} \setminus \Delta_i$, we must estimate the number of columns c of L' for which $\sigma(2, c, 2)$ does not hit row 1. When dealing with general n, m and i with $t_i = (r, c)$, to determine the number of switches from a rectangle $L' \in \Delta_i$ to a rectangle in $\Delta_{i-1} \setminus \Delta_i$, we will estimate the number of columns c' of L' for which $\sigma(c, c', r)$ does not hit any row $r' < r$. This seems like a difficult task in general. However, when $r = 1$ this difficulty does not arise, and it is clear that we can switch L' on $\sigma(c, c', 1)$ for any choice of $c' \in [n] \setminus [c]$ such that $L'_{r,c'} \notin [m]$ and obtain a rectangle in $\Delta_{i-1} \setminus \Delta_i$. This explains why we can provide non-trivial bounds on (3.2) when t_i is in the first row. By using symbol cycle and incomplete symbol cycle switching instead of column cycle and incomplete column cycle switching and applying the same logic, we see why we can obtain non-trivial bounds on (3.2) when t_i is in the first column. Note that we could also use row cycle switching to achieve this, but we get stronger bounds when we use symbol cycle and incomplete symbol cycle switching. When $t_i = (r, m)$ is in the m-th column and $c \in [n] \setminus [m]$ we can estimate the number of rectangles in Δ_i for which $\sigma(m, c, r)$ does not hit any row $r' < r$ (see Lemma 3.4). We can obtain similar results for when t_i is in the m-th row. This allows us to provide non-trivial bounds on (3.2) in these cases.

The following lemma allows us to bound (3.2) when t_i is in the first row or column.

Lemma 3.2. *Let $i \in \{m+1-\alpha, m^2+1-\alpha m\}$ for some $\alpha \in [m]$. For a random $k \times n$ Latin rectangle* $\mathbf{L}$*,*

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) = \frac{\alpha}{n+\alpha-m}.$$

Proof. First suppose that $i = m^2+1-\alpha m$ for some $\alpha \in [m]$. We use column cycle and incomplete column cycle switching to find the relative sizes of Δ_i and Δ_{i-1}. Let $L \in \Delta_i$. Let $c' \in [n] \setminus [m-\alpha+1]$ be such that $L_{1,c'} \notin [m]$. Note that there are exactly $n-m$ choices for c'. Let L' be obtained from L by switching on $\sigma(m-\alpha+1, c', 1)$. Then $L'_{1,m-\alpha+1} \notin [m]$ and $L'_{r,c} = L_{r,c}$ for every cell $(r,c) \in T_{i-1}$. Thus $L' \in \Delta_{i-1} \setminus \Delta_i$. It follows that there are exactly $(n-m)|\Delta_i|$ switches from a rectangle in Δ_i to a rectangle in $\Delta_{i-1} \setminus \Delta_i$.

Reversing the switching process, consider $L' \in \Delta_{i-1} \setminus \Delta_i$ that is obtained from some $L \in \Delta_i$ by switching on some $\sigma(m-\alpha+1, c', 1)$. Here c' must be a column in $[n] \setminus [m-\alpha+1]$ such that $L'_{1,c'} \in [m]$. Since $L'_{1,\ell} \in [m]$ for $\ell \in [m-\alpha]$ and $L'_{1,m-\alpha+1} \notin [m]$, there are exactly α choices for c'. Thus there are exactly $\alpha|\Delta_{i-1} \setminus \Delta_i|$ switches from a rectangle in $\Delta_{i-1} \setminus \Delta_i$ to a rectangle in Δ_i.

We have shown that $(n-m)|\Delta_i| = \alpha|\Delta_{i-1} \setminus \Delta_i|$. Therefore,

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) = \frac{|\Delta_i|}{|\Delta_i| + |\Delta_{i-1} \setminus \Delta_i|} = \frac{\alpha}{n+\alpha-m},$$

as required. To deal with the case where $i = m+1-\alpha$ we use symbol cycle and incomplete symbol cycle switching instead of column cycle and incomplete column cycle switching. Crucially, no two cells in $T_{m+1-\alpha}$ contain the same symbol, since we look at the first column before we look at any other cells. $\square$

Our next task is to bound (3.2) when t_i is in the m-th row or column. To do that we need to prove the preliminary results Lemma 3.3 and Lemma 3.4 below. We also need the following definitions. For a set $T \subseteq [m]^2$ of cells define $\Gamma(T)$ to be the set of $k \times n$ partial Latin rectangles P such that the set of non-empty cells of P is exactly T and every symbol of P is in $[m]$. Also define

$$\Gamma^*(T) = \bigcup_{T' \subseteq T} \Gamma(T'). \tag{3.3}$$

Lemma 3.3. *Let $P \in \Gamma^*([m]^2)$, and let (r,c) be a non-empty cell of P. Also let $r' \in [k] \setminus [m]$. Define $C' = \{c, c_1, c_2, \ldots, c_\ell\} \subseteq [m]$ to be the set of columns c' for which cell (r,c') of P is non-empty. Suppose that the entries of P in columns c and c_i form a union of column cycles, for every $i \in [\ell]$. Let A be the set of $k \times n$ Latin rectangles containing P, and let $X \subseteq A$ be the set of rectangles in A such that the row cycle $\rho(r,r',c)$ does not hit any column in $C' \setminus \{c\}$. Let* $\mathbf{L}$ *be a random $k \times n$ Latin rectangle. Then*

$$\Pr(\mathbf{L} \in X \mid \mathbf{L} \in A) = \frac{1}{\ell+1}.$$

Proof. Let $U \subseteq [m]$ be the set of rows containing a non-empty cell of P in column c. Define $Y = A \setminus X$ and a map $\phi : Y \to X$ as follows. Let $L \in Y$ and let $s = L_{r,c}$. Write the cycle of $\tau_{r,r'}$ containing s as $(s, x_1, x_2, \ldots, x_u)$. Let $v \leqslant u$ be maximal such that $x_v = L_{r,c_j}$ for some $j \in [\ell]$ and let $c' = c_j$ for this particular j. Note that v exists since $L \in Y$. Also note that $v < u$ since if $v = u$ this would imply that $(r', c', s) \in L$, contradicting the fact that the entries of P in columns c and c' form a union of column cycles. Define $\phi(L)$ by swapping the symbols in cells (r'', c) and (r'', c') for every $r'' \in [k] \setminus U$. The fact that the entries of P in columns c and c' form a union of column cycles implies that $\phi(L)$ is reached from L by switching one or more column cycles or incomplete column cycles. Hence, $\phi(L)$ is indeed a Latin rectangle and $\phi(L) \in A$. The cycle of $\tau_{r,r'}$ of $\phi(L)$ containing s is $(s, x_{v+1}, \ldots, x_u)$, meaning that $\phi(L) \in X$.

Let $L' \in X$. We now argue that there are exactly ℓ rectangles $L \in Y$ such that $\phi(L) = L'$. Note that L' is obtained from any such L by swapping all symbols in cells (r'', c) and (r'', c_w) for some $w \in [\ell]$, and all $r'' \in [k] \setminus U$. It is immediate that there are at most ℓ possible rectangles L such that $\phi(L) = L'$. We need to show that each of the ℓ possibilities is realised. The fact that $L' \in X$ ensures that $\tau_{r,r'}$ of L' contains two separate cycles $(\alpha_1, \alpha_2, \ldots, \alpha_a)$ and $(\beta_1, \beta_2, \ldots, \beta_b)$ where $\alpha_1 = L'_{r,c}$ and $\beta_1 = L'_{r,c_w}$. Swapping the contents of cells (r'', c) and (r'', c_w) for all $r'' \in [k] \setminus U$ produces a Latin rectangle in which $\tau_{r,r'}$ contains the cycle $(\alpha_1, \beta_2, \ldots, \beta_b, \beta_1, \alpha_2, \ldots, \alpha_a)$, which hits both columns c and c_w. By definition such a rectangle is in Y. It follows that $|Y| = \ell|X|$ and

$$\Pr(\mathbf{L} \in X \mid \mathbf{L} \in A) = \frac{|X|}{|X| + |Y|} = \frac{1}{\ell + 1},$$

as required. $\square$

Lemma 3.4. *Let $P \in \Gamma^*([m]^2)$, and let (r, c) be a non-empty cell of P. Also let $c' \in [n] \setminus [m]$. Define $R' = \{r, r_1, r_2, \ldots, r_\ell\} \subseteq [m]$ to be the set of rows r' for which cell (r', c) of P is non-empty. Suppose that the entries of P in rows r and r_i form a union of row cycles, for every $i \in [\ell]$. Let A be the set of $k \times n$ Latin rectangles containing P, and let $X \subseteq A$ be the set of rectangles in A such that $\sigma(c, c', r)$ does not hit any row in $R' \setminus \{r\}$. Let $\mathbf{L}$ be a random $k \times n$ Latin rectangle Then*

$$\Pr(\mathbf{L} \in X \mid \mathbf{L} \in A) \geqslant \frac{1}{\ell + 1}.$$

Proof. The proof is similar to the proof of Lemma 3.3. Let $U \subseteq [m]$ be the set of columns containing a non-empty cell of P in row r. Define $Y = A \setminus X$ and a map $\phi : Y \to X$ as follows. Let $L \in Y$ and let $s = L_{r,c}$. First suppose that $\sigma(c, c', r)$ is a column cycle. Write the cycle of $\kappa_{c,c'}$ containing s as $(s, x_1, x_2, \ldots, x_u)$. Let $v < u$ be maximal such that $x_v = L_{r_j,c}$ for some $j \in [\ell]$ and let $r' = r_j$ for this particular j. By the same arguments as in the proof of Lemma 3.3 we know that v exists. Define $\phi(L)$ by swapping the symbols in cells (r, c'') and (r', c'') for every $c \in [n] \setminus U$. The fact that the entries of P in rows r and r' form a union of row cycles implies that $\phi(L)$ is reached from L by switching on one or more row cycles. Hence, $\phi(L)$ is indeed a Latin rectangle and $\phi(L) \in A$. The cycle of $\kappa_{c,c'}$ of $\phi(L)$ containing symbol s is $(s, x_{v+1}, \ldots, x_u)$ meaning that $\phi(L) \in X$. Now suppose that $\sigma(c, c', r)$ is an incomplete column cycle. Write the incomplete cycle of $\kappa_{c,c'}$ as $[y_1, \ldots, y_{u'}, s, x_1, \ldots, x_u]$. We now consider two cases. First, suppose that there is some $v \in [u']$ such that $L_{r_j,c} = y_v$ for some $j \in [\ell]$. Let such a v be maximal and let $r' = r_j$ for this value of j. If $v = u'$ then L must contain the entry (r', c', s), contradicting the fact that the entries of P in rows r and r' form a union of row cycles. So $v < u'$. Let $\phi(L)$ be obtained from L by swapping the contents in cells (r, c'') and (r', c'') for every $c'' \in [n] \setminus U$. The fact that the entries of P in rows r and r' form a union of row cycles implies that $\phi(L)$ is reached from L by switching one or more row cycles. Hence, $\phi(L)$ is indeed a Latin rectangle and $\phi(L) \in A$. In $\phi(L)$, $\sigma(c, c', r)$ is a cycle and the cycle of $\kappa_{c,c'}$ of $\phi(L)$ containing s is $(s, y_{v+1}, \ldots, y_{u'})$. Thus, $\phi(L) \in X$. Finally, suppose that no such v exists. Since $L \in Y$ there is some $v' < u$ such that $L_{r_j,c} = x_{v'}$. Let such a v' be maximal and let $r' = r_j$ for this value of j. Let $\phi(L)$ be obtained from L by swapping the contents in cells (r, c'') and (r', c'') for every $c'' \in [n] \setminus U$. Then $\phi(L) \in A$ and the incomplete cycle of $\kappa_{c,c'}$ containing s is $[y_1, \ldots, y_{u'}, s, x_{v'+1}, \ldots, x_u]$. Therefore, $\phi(L) \in X$.

Let $L' \in X$. We now argue that there are at most ℓ rectangles $L \in Y$ such that $\phi(L) = L'$. This is true since L' is obtained from any such L by swapping all symbols in cells (r, c'') and (r_w, c'') for some $w \in [\ell]$, and all $c'' \in [n] \setminus U$. It follows that $|Y| \leqslant \ell |X|$ and

$$\Pr(\mathbf{L} \in X \mid \mathbf{L} \in A) = \frac{|X|}{|X| + |Y|} \geqslant \frac{1}{\ell + 1},$$

as required. □

We can now bound (3.2) when t_i is in the m-th row.

Lemma 3.5. *Let $i = jm$ where $j \in [m]$ and let $\mathbf{L}$ be a random $k \times n$ Latin rectangle. Then*

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) \leqslant \frac{j}{k + j - m}.$$

Proof. We use row cycle switching to estimate the relative sizes of Δ_i and $\Delta_{i-1} \setminus \Delta_i$. For a partial Latin rectangle $P \in \Gamma(T_i)$ let $\mathcal{L}_P$ denote the set of $k \times n$ Latin rectangles containing P. So Δ_i is the disjoint union

$$\bigcup_{P \in \Gamma(T_i)} \mathcal{L}_P.$$

Let $r' \in [k] \setminus [m]$. Note that if $L \in \Delta_i$ then $L_{m,j} \in [m]$ and $L_{r',j} \notin [m]$. Let $P \in \Gamma(T_i)$. By Lemma 3.3 there are $|\mathcal{L}_P|/j$ rectangles $L \in \mathcal{L}_P$ such that switching L on $\rho(m, r', j)$ yields a rectangle in $\Delta_{i-1} \setminus \Delta_i$. Thus the number of switches from a rectangle in Δ_i to one in $\Delta_{i-1} \setminus \Delta_i$ is $(k - m)|\Delta_i|/j$.

Reversing the switching, consider $L' \in \Delta_{i-1} \setminus \Delta_i$. Note that $L'_{m,j} \notin [m]$ and there is at most one $r' \in [k] \setminus [m]$ such that $L'_{r',j} \in [m]$. Therefore there is at most one switch from L' to a rectangle in Δ_i.

We have thus shown that $(k - m)|\Delta_i|/j \leqslant |\Delta_{i-1} \setminus \Delta_i|$ and so

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) = \frac{|\Delta_i|}{|\Delta_i| + |\Delta_{i-1} \setminus \Delta_i|} \leqslant \frac{j}{k + j - m},$$

as required. □

Using column cycle and incomplete column cycle switching instead of row cycle switching we can bound (3.2) when t_i is in the m-th column.

Lemma 3.6. *Let $i = m^2 - m + j$ where $j \in [m]$ and let $\mathbf{L}$ be a random $k \times n$ Latin rectangle. Then*

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) \leqslant \frac{j}{n + j - m}.$$

Proof. The proof is similar to the proof of Lemma 3.5. We will again denote the set of $k \times n$ Latin rectangles containing P by $\mathcal{L}_P$ for $P \in \Gamma(T_i)$.

Let $c' \in [n] \setminus [m]$. Note that if $L \in \Delta_i$ then $L_{j,m} \in [m]$ and $L_{j,c'} \notin [m]$. Let $P \in \Gamma(T_i)$. By Lemma 3.4 there are at least $|\mathcal{L}_P|/j$ rectangles $L \in \mathcal{L}_P$ such that switching L on $\sigma(m, c', j)$ yields a rectangle in $\Delta_{i-1} \setminus \Delta_i$. Thus the number of switches from a rectangle in Δ_i to one in $\Delta_{i-1} \setminus \Delta_i$ is at least $(n - m)|\Delta_i|/j$.

Reversing the switching, consider $L' \in \Delta_{i-1} \setminus \Delta_i$. Note that $L'_{j,m} \notin [m]$ and there is a unique $c' \in [n] \setminus [m]$ such that $L'_{j,c'} \in [m]$. Therefore there is at most one switch from L' to a rectangle in Δ_i, namely $\sigma(m, c', j)$.

We have thus shown that $(n-m)|\Delta_i|/j \leqslant |\Delta_{i-1} \setminus \Delta_i|$ and so

$$\Pr(\mathbf{L} \in \Delta_i \mid \mathbf{L} \in \Delta_{i-1}) = \frac{|\Delta_i|}{|\Delta_i| + |\Delta_{i-1} \setminus \Delta_i|} \leqslant \frac{j}{n+j-m},$$

as required. $\square$

We are now ready to prove Theorem 3.1.

Proof of Theorem 3.1. As previously mentioned, the probability that $\mathbf{L}[R, C]$ is a Latin square of order m with symbol set S is equal to the probability that $\mathbf{L} \in \Delta_{m^2}$. By (3.1) combined with Lemma 3.2, Lemma 3.5 and Lemma 3.6 we obtain the bound

$$\begin{aligned}\Pr(\mathbf{L} \in \Delta_{m^2}) &\leqslant \left(\prod_{j=1}^{m} \frac{j}{n+j-m}\right)^2 \left(\prod_{j=2}^{m-1} \frac{j}{n+j-m}\right) \left(\prod_{j=2}^{m-1} \frac{j}{k+j-m}\right) \\ &= \frac{nk(n+1-m)(k+1-m)}{m^2 \binom{n}{m}^3 \binom{k}{m}}\end{aligned}$$

as required. $\square$

3.2 Bounding probability of small subsquares

The aim of this subsection is to prove the following theorem, which will enable us to bound the expected number of small subsquares. Our bound improves the bound in the previous subsection when $2 < m = O(1)$, although it is worse when m grows at least logarithmically in n. The new bound will be used when $m \in \{4, 5\}$ to get the accuracy required for Theorem 1.1.

Theorem 3.7. *Let $R \subseteq [k]$ be of cardinality m, let C and S be subsets of $[n]$ of cardinality m, and let $\mathbf{L}$ be a random $k \times n$ Latin rectangle where $k > m > 2$. The probability that $\mathbf{L}[R, C]$ is a Latin square of order m with symbol set S is at most*

$$\frac{2^{m-1} m^{m^2+2m-5}}{(k-m)^{5m/2-6}} \left(\frac{1+O(m/n)}{n}\right)^{3m-2}$$

if m is even and at most

$$\frac{2^{m-2} m^{m^2+2m-5}}{(k-m)^{(5m-13)/2}} \left(\frac{1+O(m/n)}{n}\right)^{3m-2}$$

if m is odd. If $k = m > 2$ then the same bounds hold with $k - m$ replaced by 1.

As in §3.1, we may assume that $R = C = S = [m]$. Also, if M and M' are any two Latin squares of order m, then replacing M by M' gives an easy bijection from Latin rectangles containing M as a subsquare to Latin rectangles containing M'. Hence, if $\mathbf{L}$ is a random $k \times n$ Latin rectangle then

$$\Pr(\mathbf{L}[[m],[m]] = M) = \Pr(\mathbf{L}[[m],[m]] = M'). \tag{3.4}$$

Thus without loss of generality we may assume that $\mathbf{L}[[m],[m]] = M$ is a specific Latin square of order m, which we will choose later. Ultimately, we will multiply by m^{m^2}, which is a trivial upper bound on the number of Latin squares of order m, to obtain Theorem 3.7. The strategy is similar to the proof of Theorem 3.1. Recalling (3.3), we let P and P' be elements of $\Gamma^*([m]^2)$ such that $P \subseteq P'$ and $P' \setminus P = \{(r,c,s)\}$. We will give non-trivial upper bounds on

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \tag{3.5}$$

when P and P' satisfy certain conditions.

First, we bound (3.5) when (r,c,s) is the only occurrence of row r, the only occurrence of column c, or the only occurrence of symbol s in P'.

Lemma 3.8. *Let $\{P,P'\} \subseteq \Gamma^*([m]^2)$ be such that $P \subseteq P'$ and $P' \setminus P = \{(r,c,s)\}$. Let $\mathbf{L}$ be a random $k \times n$ Latin rectangle, where $k > m$. Suppose that P has no entry in row r. Then*

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \leqslant \frac{1}{k-m}.$$

If P has no entry in column c or no entry with symbol s then

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \leqslant \frac{1+O(m/n)}{n}.$$

Proof. First consider the case where P has no entry in row r. Let Y be the set of $k \times n$ Latin rectangles containing P and let $X \subseteq Y$ be the set of rectangles containing P'. We use row cycle switching to find bounds on the relative sizes of X and $Y \setminus X$.

Let $L \in X$. For every row $r' \in [k] \setminus [m]$, switching on $\rho(r,r',c)$ yields a rectangle in $Y \setminus X$. This gives exactly $k-m$ switches from L to a rectangle in $Y \setminus X$.

Now, consider $L' \in Y \setminus X$ and note that there is at most one $r' \in [k]$ such that $L'_{r',c} = s$. Thus there is at most one row cycle which can be switched back to change L' into some rectangle in X. Thus at most one switch counts from L' to a rectangle in X.

We have deduced that $(k-m)|X| \leqslant |Y \setminus X| \leqslant |Y|$. Thus

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) = \frac{|X|}{|Y|} \leqslant \frac{1}{k-m}.$$

To deal with the case where P has no entry in column c or has no entry with symbol s we use analogous arguments with column cycle and incomplete column cycle switching or symbol cycle and incomplete column cycle switching, respectively. In each case there are exactly $n-m$ switches from L to a rectangle in $Y \setminus X$ and hence

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \leqslant \frac{1}{n-m} = \frac{1+O(m/n)}{n}. \qquad \square$$

We now bound (3.5) when P' contains a cell (r,c) satisfying the conditions of Lemma 3.3.

Lemma 3.9. *Let $\{P,P'\} \subseteq \Gamma^*([m]^2)$ be such that $P \subseteq P'$ and $P' \setminus P = \{(r,c,s)\}$. Let C' be the set of columns c' for which the cell (r,c') of P' is non-empty. Suppose further that for every $c' \in C' \setminus \{c\}$, the entries of P' in columns c and c' form a union of column cycles. Let $\mathbf{L}$ be a random $k \times n$ Latin rectangle where $k > m$. Then*

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \leqslant \frac{|C'|}{k-m}.$$

Proof. Let Y be the set of $k \times n$ Latin rectangles containing P and let $X \subseteq Y$ be the set of rectangles containing P'.

Let $r' \in [k] \setminus [m]$. By Lemma 3.3 there are $|X|/|C'|$ rectangles in X for which switching on $\rho(r, r', c)$ yields a rectangle in $Y \setminus X$. Thus there are $(k-m)|X|/|C'|$ switches from X to $Y \setminus X$.

Now let $L' \in Y \setminus X$. The same argument as given in the proof of Lemma 3.8 tells us that there is at most one row cycle that can be switched from L' to reach a rectangle in X.

We have shown that $(k-m)|X|/|C'| \leqslant |Y \setminus X| \leqslant |Y|$. Thus

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) = \frac{|X|}{|Y|} \leqslant \frac{|C'|}{k-m},$$

as required. □

By using column cycle and incomplete column cycle switching rather than row cycle switching and employing Lemma 3.4 we can prove the following lemma.

Lemma 3.10. *Let $\{P, P'\} \subseteq \Gamma^*([m]^2)$ be such that $P \subseteq P'$ and $P' \setminus P = \{(r, c, s)\}$. Let R' be the set of rows r' for which the cell (r', c) of P' is non-empty. Suppose further that for every $r' \in R' \setminus \{r\}$, the entries of P' in rows r and r' form a union of row cycles. Let $\mathbf{L}$ be a random $k \times n$ Latin rectangle. Then*

$$\Pr(\mathbf{L} \supseteq P' \mid \mathbf{L} \supseteq P) \leqslant \frac{|R'|}{n}(1 + O(m/n)).$$

Our next task is to choose the Latin square M of order m. Our choice of M allows us to apply Lemma 3.8, Lemma 3.9 or Lemma 3.10 many times to give non-trivial bounds on (3.5). If m is even then define M to be a Latin square of order m with the following properties:

- $M_{1,i} = M_{i,1} = i$ for every $i \in [m]$,
- $M_{2,i} = M_{i,2} = i + 1$ for every odd $i \in [m]$,
- $M_{2,i} = M_{i,2} = i - 1$ for every even $i \in [m]$.

If m is odd then define M to be a Latin square of order m satisfying:

- $M_{1,i} = M_{i,1} = i$ for every $i \in [m]$,
- $M_{2,i} = M_{i,2} = i + 1$ for every odd $i \in [m-3]$,
- $M_{2,i} = M_{i,2} = i - 1$ for every even $i \in [m-3]$,
- $(M_{2,m-2}, M_{2,m-1}, M_{2,m}) = (M_{m-2,2}, M_{m-1,2}, M_{m,2}) = (m-1, m, m-2)$.

Such a square M exists by [1, Theorem 1.5]. See Figure 2 for examples of such squares when $m = 4$ and $m = 5$. These are the only two orders for which Theorem 3.7 will be used in the proof of Theorem 1.1.

We are now ready to prove Theorem 3.7.

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 4 & 3 \\ 3 & 4 & 1 & 2 \\ 4 & 3 & 2 & 1 \end{bmatrix} \qquad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \\ 3 & 4 & 5 & 1 & 2 \\ 4 & 5 & 2 & 3 & 1 \\ 5 & 3 & 1 & 2 & 4 \end{bmatrix}$$

Figure 2: The Latin squares M when $m \in \{4, 5\}$. Colour coding , , indicates cells to which Lemmas 3.8, 3.9 and 3.10, respectively, will be applied in the proof of Theorem 3.7.

Proof of Theorem 3.7. First suppose that m is even and $k > m$. We assume that $R = C = S = [m]$ and that M is the Latin square of order m defined above. Let P be the $k \times n$ partial Latin rectangle such that $P[[m], [m]] = M$ and every cell of P outside of $[m]^2$ is empty. We define a family of partial Latin rectangles $P_i \subseteq P$ as follows. Firstly, $P_0 = \varnothing$. Then for $i \in [m^2]$ we define P_i by adding an entry of P to P_{i-1}.

For $i \in [2m]$ we add the entry of P in cell $(1, (i+1)/2)$ if i is odd and we add the entry in cell $(2, i/2)$ if i is even. If i is odd then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant (1 + O(m/n))/n \tag{3.6}$$

by Lemma 3.8, since there is only one non-empty cell of P_i in column $(i+1)/2$. Similarly, if $i \equiv 2 \bmod 4$ then (3.6) holds by Lemma 3.8, since the entry $(2, i/2, i/2+1)$ is the only entry of P_i with symbol $i/2 + 1$. If $i \equiv 0 \bmod 4$ then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant 2(1 + O(m/n))/n, \tag{3.7}$$

by Lemma 3.10.

For $i \in \{2m+1, \dots, 4m-4\}$ we add the entry of P in cell $((i-2m+5)/2, 1)$ if i is odd and we add the entry in cell $((i-2m+4)/2, 2)$ if i is even. If i is odd then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant 1/(k-m) \tag{3.8}$$

holds by Lemma 3.8, since there is only one non-empty cell of P_i in row $(i-2m+5)/2$. If $i \equiv 0 \bmod 4$ then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant 2/(k-m) \tag{3.9}$$

holds by Lemma 3.9. If $i \equiv 2 \bmod 4$ we simply use $\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant 1$.

For $i \in \{4m-3, \dots, m^2\}$ we add the entry of P in cell $(3 + \lfloor (i-4m+3)/(m-2) \rfloor, 3 + ((i-4m+3) \bmod m-2))$ (this is the natural ordering for the cells in $\{3, 4, \dots, m\}^2$). If $i \equiv 4 \bmod m-2$ (in which case the added entry is in column m) then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant m(1 + O(m/n))/n, \tag{3.10}$$

by Lemma 3.10. If $i \geqslant m^2 - m + 3$ (in which case the added entry is in row m) then

$$\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant m/(k-m) \tag{3.11}$$

by Lemma 3.9. For all other values of i we use $\Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \leqslant 1$.

Overall, we use (3.6) for $3m/2$ values of i, corresponding to the m entries in the first row and $m/2$ entries in the second row. We use (3.7) for $m/2$ values of i, corresponding to entries in the second row. We use (3.8) for $m-2$ values of i, corresponding to entries in the first column. We use (3.9) for $m/2-1$ values of i, corresponding to entries in the second column. We use (3.10) for $m-2$ values of i, corresponding to entries in the m-th column. We use (3.11) for $m-3$ values of i, corresponding to entries in the m-th row. Hence, by the chain rule of probability,

$$\begin{aligned}\Pr(\mathbf{L} \supseteq P) &= \prod_{i=1}^{m^2} \Pr(\mathbf{L} \supseteq P_i \mid \mathbf{L} \supseteq P_{i-1}) \\ &\leqslant \left(\frac{1+O(m/n)}{n}\right)^{3m/2} \left(\frac{2(1+O(m/n))}{n}\right)^{m/2} \left(\frac{1}{k-m}\right)^{m-2} \qquad (3.12) \\ &\quad \cdot \left(\frac{2}{k-m}\right)^{m/2-1} \left(\frac{m(1+O(m/n))}{n}\right)^{m-2} \left(\frac{m}{k-m}\right)^{m-3} \\ &= \frac{2^{m-1}m^{2m-5}}{(k-m)^{5m/2-6}} \left(\frac{1+O(m/n)}{n}\right)^{3m-2}. \qquad (3.13)\end{aligned}$$

If m is even and $k = m$ then the same bounds hold except we use the trivial bound in place of (3.8), (3.9) and (3.11).

The argument when m is odd is similar. We use (3.6) for $(3m+1)/2$ entries, namely m entries in the first row and $(m+1)/2$ entries in the second row. We use (3.7) for $(m-1)/2$ entries in the second row. If $k > m$ then we use (3.8) for $m-2$ entries in the first column and (3.9) for $(m-3)/2$ entries in the second column. We use (3.10) for $m-2$ entries in the m-th column and if $k > m$ then we use (3.11) for $m-3$ entries in the m-th row. As a consequence,

$$\Pr(\mathbf{L} \supseteq P) \leqslant \frac{2^{m-2}m^{2m-5}}{(k-m)^{(5m-13)/2}} \left(\frac{1+O(m/n)}{n}\right)^{3m-2}, \qquad (3.14)$$

with $k-m$ replaced by 1 in the case when $k = m$. As foreshadowed after (3.4), we now multiply by m^{m^2} to give the claimed upper bound on $\mathbf{L}[R, C]$ being any subsquare on the symbols $[m]$. $\square$

3.3 Proof of Theorem 1.1

We are now ready to prove our main theorem.

Proof of Theorem 1.1. Let m be an integer satisfying $4 \leqslant m \leqslant \min\{k, n/2\}$. There are $\binom{k}{m}$ choices for the set R and there are $\binom{n}{m}^2$ choices for the sets C and S in Theorem 3.1 and Theorem 3.7. From Theorem 3.1 and $\binom{n}{m} \geqslant (n/m)^m$, we obtain

$$\mathbb{E}_m(k,n) \leqslant \frac{kn(k+1-m)(n+1-m)}{m^2\binom{n}{m}} \leqslant \frac{m^{m-2}}{n^{m-4}} = \frac{m^4}{n^2}\left(\frac{m}{n}\right)^{m-6} \leqslant \frac{m^4}{n^2 2^{m-6}} = O(n^{-2}), \qquad (3.15)$$

provided $m \geqslant 6$.

For $m \in \{4, 5\}$ and $k > m$ we use Theorem 3.7 to deduce that

$$\mathbb{E}_4(k,n) \leqslant \binom{k}{4}\binom{n}{4}^2 O(1) \left(\frac{1+O(1/n)}{n}\right)^{10} \left(\frac{1}{k-4}\right)^4 = O(n^{-2}),$$

and

$$\mathbb{E}_5(k,n) \leqslant \binom{k}{5}\binom{n}{5}^2 O(1) \left(\frac{1+O(1/n)}{n}\right)^{13} \left(\frac{1}{k-5}\right)^6 = O(n^{-3}).$$

When $k = m \in \{4, 5\}$ all terms involving k can be omitted, with the conclusion unchanged. Combining with (3.15), we have the result. $\square$

4 Subsquares of order 2 or 3

From Theorem 1.1 we know that $\mathbb{E}_m(k,n)$ is asymptotically 0 for all $m \geqslant 4$. In this last section we consider subsquares of order $m < 4$. McKay and Wanless [9] conjectured that $\mathbb{E}_3(n,n) = 1/18 + o(1)$. Kwan, Sah and Sawhney [6] conjectured further that the distribution of the number of 3×3 subsquares in a random Latin square is asymptotically Poisson with this mean. If these conjectures are true, then $\lim_{n\to\infty} \mathbb{E}_m(n,n)$ is both positive and finite only when $m = 3$. Divoux, Kelly, Kennedy and Sidhu [3] generalised McKay and Wanless' conjecture by suggesting that $\mathbb{E}_3(k,n) = (1+o(1))\binom{k}{3}/(3n^3)$. Although we are not able to resolve the situation completely, we are able to prove that $\mathbb{E}_3(k,n)$ is bounded. The proof follows the derivation of (3.14) closely, but when $m = 3$ that result was slightly conservative when counting entries in the second column to which Lemma 3.9 can be applied.

Theorem 4.1. *For $3 \leqslant k \leqslant n$,*

$$\mathbb{E}_3(k,n) \leqslant \frac{2k}{3n}(1 + O(1/k)).$$

Proof. Let M denote the Latin square

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \\ 3 & 1 & 2 \end{bmatrix},$$

with the same colouring convention as used in Figure 2. Let P be the partial $k \times n$ Latin rectangle such that $P[[3],[3]] = M$ and every cell of P outside of $[3]^2$ is empty. We first bound the probability that a random $k \times n$ Latin rectangle $\mathbf{L}$ contains P in the case when $k > 3$. We define a family of partial Latin rectangles $P_i \subseteq P$ as follows. Let $P_0 = \varnothing$ and for $i \in [9]$ let P_i be obtained from P_{i-1} by adding the entry of P in the cell given by the i-th element of the tuple $((1,1),(2,1),(1,2),(2,2),(1,3),(2,3),(3,1),(3,2),(3,3))$. For $i \in [5]$ we have that (3.6) holds by Lemma 3.8. For $i = 6$ we have that (3.7) holds by Lemma 3.10. For $i = 7$ we have that (3.8) holds by Lemma 3.8. For $i = 8$ we have that (3.9) holds by Lemma 3.9. Finally, we see that $\Pr(\mathbf{L} \supseteq P_9 \mid \mathbf{L} \supseteq P_8) \leqslant 3(1 + O(1/n))/n$ by Lemma 3.10. The chain rule of probability implies that

$$\Pr(\mathbf{L} \supseteq P) \leqslant 12 \left(\frac{1 + O(1/n)}{n}\right)^7 \left(\frac{1}{k-3}\right)^2 = \frac{12}{n^7(k-3)^2}(1 + O(1/n)).$$

Since there are 12 Latin squares of order 3 we obtain

$$\mathbb{E}_3(k,n) \leqslant \binom{k}{3}\binom{n}{3}^2 \frac{144}{n^7(k-3)^2}(1 + O(1/n)) = \frac{2k}{3n}(1 + O(1/k)),$$

as required. If $k = 3$ then we instead get

$$\mathbb{E}_3(k,n) \leqslant \binom{n}{3}^2 \frac{72}{n^7}(1 + O(1/n)) = \frac{2}{n}(1 + O(1/n)). \qquad \square$$

Finally, we turn to the case $m = 2$. Recall from Theorem 2.1 that $\mathbb{E}_2(n,n) = (1 + o(1))n^2/4$. Also, [3, Corollary 1.7] tells us that $\mathbb{E}_2(k,n) = (1/2 + o(1))\binom{k}{2}$ whenever $k \leqslant (1/2 - o(1))n$. We extend this result to all $k \leqslant n$. We first need the following lemma.

Lemma 4.2. *Let* $\mathbf{L}$ *be a random* $k \times n$ *Latin rectangle and let* $\{i,j\} \subseteq [k]$ *with* $i \neq j$. *The probability that* $\mathbf{L}[\{i,j\},[n]]$ *contains at least* t *intercalates is at most* $\exp(-\Omega(t \log t))$.

Proof. Let P be the partial $k \times n$ Latin rectangle which is empty except that:

- $P_{i,\ell} = \ell$ for every $\ell \in [2t]$,
- $P_{j,\ell} = \ell + 1$ for every odd $\ell \in [2t]$, and
- $P_{j,\ell} = \ell - 1$ for every even $\ell \in [2t]$.

First suppose that $t \leqslant n/4$, so that we can use the argument behind (3.6) and (3.7) to obtain

$$\Pr(\mathbf{L} \supseteq P) \leqslant 2^t n^{-4t}(1 + O(t/n))^{4t}.$$

For t intercalates in general position in $\mathbf{L}[\{i,j\},[n]]$ there are $\binom{n}{2t}$ choices for the symbols, $(2t)!/(2^t t!)$ ways to pair them, and then $n!/(n-2t)!$ ways to allocate the columns. Thus the probability that $\mathbf{L}[\{i,j\},[n]]$ contains at least t intercalates is at most

$$\left(\frac{n!}{(n-2t)!}\right)^2 \frac{1}{2^t t!} 2^t n^{-4t}(1 + O(t/n))^{4t} = \exp(-\Omega(t \log t)),$$

as required. If $n/4 < t \leqslant n/2$ then $\Omega(t \log t) = \Omega((n/4)\log(n/4))$, so we can derive the same bound simply by considering the first $n/4$ intercalates in $\mathbf{L}[\{i,j\},[n]]$. □

Theorem 4.3. *For* $2 \leqslant k \leqslant n$,

$$\mathbb{E}_2(k,n) = \frac{1}{2}\binom{k}{2}(1 + o(1)).$$

Proof. Since [3, Corollary 1.7] proves the claim when $k \leqslant (1/2 - o(1))n$, we may assume that $k = \Theta(n)$, which means that arguments that have previously been used for counting intercalates in Latin squares apply with only minor changes. Let k' be a integer function of n with $0 \leqslant k' \leqslant k$. For a $k' \times n$ Latin rectangle T let $\mathcal{R}(T)$ be the set of $k \times n$ Latin rectangles L such that $L[[k'],[n]] = T$. By [9, Proposition 4], if T and T' are $k' \times n$ Latin rectangles, then

$$\frac{|\mathcal{R}(T)|}{|\mathcal{R}(T')|} \leqslant \exp(O(n \log^2 n)). \tag{4.1}$$

Let $\mathbf{N}$ denote the number of intercalates in a random $k \times n$ Latin rectangle. Modifying the proof of [8, Theorem 4], by using Lemma 4.2 instead of [8, Theorem 3] and using (4.1), gives

$$\Pr(\mathbf{N} \leqslant (1-\delta)k^2/4) \leqslant \exp(-\Omega(n^{1/2}\log^{-1} n)), \tag{4.2}$$

for fixed $\delta \in (0,1]$. Similarly, a simple adaptation of the proof of [7, Theorem 2.1] using (4.1) gives

$$\Pr(\mathbf{N} \geqslant (1+\delta)k^2/4) \leqslant \exp(-\Omega(n^{4/3}\log n)), \tag{4.3}$$

for fixed $\delta > 0$. Combining (4.2) and (4.3) proves the theorem. □

References

[1] P. Adams, D. Bryant, and M. Buchanan. "Completing partial Latin squares with two filled rows and two filled columns". *Electron. J. Combin.* 15.1 (2008), 26pp.

[2] N. J. Cavenagh, C. Greenhill, and I. M. Wanless. "The cycle structure of two rows in a random Latin square". *Random Structures Algorithms* 33.3 (2008), pp. 286–309.

[3] A. Divoux, T. Kelly, C. Kennedy, and J. Sidhu. "Subsquares in random Latin squares and rectangles". *arXiv:2311.04152* (2023).

[4] M. J. Gill, A. Mammoliti, and I. M. Wanless. "Canonical labelling of Latin squares in average-case polynomial time". *Random Structures Algorithms* (to appear).

[5] C. D. Godsil and B. D. McKay. "Asymptotic enumeration of Latin rectangles". *J. Combin. Theory Ser. B* 48.1 (1990), pp. 19–44.

[6] M. Kwan, A. Sah, and M. Sawhney. "Large deviations in random Latin squares". *Bull. Lond. Math. Soc.* 54.4 (2022), pp. 1420–1438.

[7] M. Kwan, A. Sah, M. Sawhney, and M. Simkin. "Substructures in Latin squares". *Israel J. Math.* 256.2 (2023), pp. 363–416.

[8] M. Kwan and B. Sudakov. "Intercalates and discrepancy in random Latin squares". *Random Structures Algorithms* 52.2 (2018), pp. 181–196.

[9] B. D. McKay and I. M. Wanless. "Most Latin squares have many subsquares". *J. Combin. Theory Ser. A* 86.2 (1999), pp. 322–347.

[10] I. M. Wanless. "Cycle switches in Latin squares". *Graphs Combin.* 20.4 (2004), pp. 545–570.