ESTIMATING THE NUMBER OF LATIN RECTANGLES BY THE FAST SIMULATION METHOD

N. Yu. Kuznetsov

A fast simulation method is proposed to estimate the number of Latin rectangles and squares. Numerous examples demonstrate the high accuracy of the method. The number of Latin squares of order \( n = 20 \) is estimated with a relative error of 5\% and a confidence level of 0.99. Statistical lower bounds for the maximum number of transversals over all Latin squares of order \( n \leq 20 \) are obtained.

Keywords: Latin rectangle, Latin square, transversal, fast simulation method, unbiased estimate, sample variance, relative error.

A Latin rectangular is a rectangular \( n \times m \) matrix, \( m \leq n \), with each column being a permutation (without repetition) of elements of a set \( S = \{ a_1, \ldots, a_n \} \), elements not repeating in each row. If \( m = n \), then we obtain a square \( n \times n \) matrix, with each row and each column being a permutation of elements of the set \( S \). A square matrix possessing this property is called a Latin square. The term “Latin” originates from Euler, who used letters of the Latin alphabet to denote elements of the set \( S \). Now, the set \( S \) is usually \( S = \{1, \ldots, n\} \).

The problem of enumerating all Latin rectangles is \( NP \)-complete and has not been solved in the general case. The problem of determining if a partially filled square can be completed to form a Latin square is also \( NP \)-complete [1]. Explicit analytical formulas to determine the number of Latin rectangles are known only for \( m = 2 \) (derangement problem) and \( m = 3 \) (ménage problem, i.e., arrangement of \( n \) married couples at a round table so that nobody sits next to his or her spouse) [2, 3]. Denote by \( L(n, m) \) the number of \( n \times m \) Latin rectangles, \( m \leq n \); \( N(n, m) \) the number of Latin rectangles with the fixed first column \( (1, 2, \ldots, n)^T \) (normalized rectangles); \( M(n, m) \) the number of Latin rectangles with the fixed first column \( (1, 2, \ldots, n)^T \) and first row \( (1, 2, \ldots, m) \) (reduced or standard rectangle).

The following obvious relations hold:

\[
L(n, m) = n! N(n, m) = n! A_{m-1}^{n-1} M(n, m) = \frac{n!(n-1)!}{(n-m)!} M(n, m).
\]  

(1)

Finding the exact number of Latin squares involves much computational efforts that exponentially increase with \( n \). The use of most advanced computers has allowed determining \( L(n, n) \) only for \( n \leq 11 \) [4]. Therefore, the main emphasis should be put on the development of approximate methods for calculating the number of Latin squares and Latin rectangles.

This paper proposes an alternative approach based on special simulation methods that allow constructing a Latin rectangle directly, with the analytical calculation of the corresponding normalizing factors whose product is the estimate. Given relative error \( \varepsilon \) and confidence probability \( \gamma \), the estimate \( \hat{L}(n, m, \varepsilon, \gamma) \) and the corresponding confidence interval \( \hat{L}(n, m, \varepsilon, \gamma) \) are found. By choosing appropriate parameters \( \varepsilon \) and \( \gamma \), we can substantially extend the domain of values of \( n \) and \( m \), such that \( L(n, m) \) can be estimated in a relatively short time. For example, such estimates are found for \( L(n, n) \) with \( n = 20 \), \( \varepsilon = 5\% \) and \( \gamma = 0.99 \) and for \( L(n, m) \) with \( n = 1000 \), \( m = 10 \), \( \varepsilon = 1\% \) and \( \gamma = 0.99 \). If so high accuracy of computations is not necessary, estimates can be found for much greater values of \( n \) and \( m \).
The method proposed is analyzed in detail for accuracy. It is checked whether exact values of $L(n, m)$ fall within the corresponding confidence intervals. Moreover, the joint behavior of statistical and asymptotic estimate is investigated.

In the last section, the statistical lower estimate of the maximum number of transversals in Latin squares is obtained and compared with recent relevant results [5].

**ESTIMATING $M(n, m)$ BY THE FAST SIMULATION METHOD**

We will pay the main attention to estimating $M(n, m)$ (the number of reduced Latin rectangles). The total number of Latin rectangles $L(n, m)$ and the number of normalized rectangles $N(n, m)$ are determined from (1). Since the first column and the first row are filled up, it remains necessary to fill in $(m-1)(n-1)$ places in the other $m-1$ columns, and a column $j$ ($2 \leq j \leq m$) can contain only the symbols $\{1, \ldots, j-1, j+1, \ldots, n\}$. Assume that they are arranged randomly ($n$! alternatives for each column). Denote by $P(n, m)$ the probability of constructing a Latin rectangle. Obviously,

$$P(n, m) = \frac{M(n, m)}{[(n-1)!]^{m-1}}. \tag{2}$$

As noted above, there are no analytical methods to find $P(n, m)$ (the only exception is the case $m \leq 3$). An alternative approach is based on the Monte Carlo method, which allows finding approximate estimates for $P(n, m)$. At the same time, as $n$ and $m$ ($m \leq n$) increase, the probability $P(n, m)$ decreases rapidly, which makes the Monte Carlo method inapplicable even for quite moderate values of $n$ and $m$.

The present section proposes the fast simulation method (an alternative method of weighted modeling), which allows selecting directly the places in which one symbol or another can be put. The unbiasedness of the estimates can be achieved by choosing appropriate weight factors. The acceleration is due to a sharp increase in the number of realizations in which it is possible to construct a Latin rectangle. The corresponding weight factors are comparable, in their order, with $P(n, m)$. This substantially decreases the variance of the estimate and thus the number of realizations necessary to achieve a required estimation accuracy. Let us introduce auxiliary indicators

$$v_i(k) = \begin{cases} 1 & \text{if } k \text{ is in the row } i, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mu_i(k) = \begin{cases} 1 & \text{if } k \text{ is in the column } i, \\ 0 & \text{otherwise} \end{cases}$$

that determine the arrangement of symbols in the rows and columns. Let us formulate the fast simulation method as an algorithm of finding the estimate $\hat{P}_f(n, m)$ in one realization for $P(n, m)$.

1. Put $q = 1$ (initial value of the normalizing factor) and specify the initial state of the matrix $A = (a_{ij})_{i=1, j=1}^{n,m}$ being filled in:

$$a_{ij}(k) = \begin{cases} i & \text{if } j = 1, \\ j & \text{if } i = 1, \\ 0 & \text{otherwise.} \end{cases}$$

In the next Steps 2–8, place symbols of the set $C = \{2, \ldots, n-1\}$. Suppose $k = 2$.

2. Specify the initial values of the auxiliary indicators:

$$v_i(k) = \begin{cases} 1 & \text{if } i = k, \\ 0 & \text{otherwise,} \end{cases} \quad \mu_j(k) = \begin{cases} 1 & \text{if } j = k, \\ 0 & \text{otherwise.} \end{cases}$$

3. Form the set $D_k = \{j: 2 \leq j \leq m, \mu_j(k) = 0\}$ of the numbers of columns where the symbol $k$ should be placed.

4. For each column, determine the number of places in which the symbol $k$ may appear:

$$r_j = \sum_{i: v_i(k) = 0, a_{ij} = 0} 1, \quad j \in D_k, \quad j_0 = \arg \min_{j \in D_k} r_j.$$