

# A Systematic Procedure for Square Random Lattice Sampling

<sup>1</sup>Girish Chandra and <sup>2</sup>Neeraj Tiwari

<sup>1</sup>Division of Forestry Statistics, Indian Council of Forestry Research and Education, Dehradun, India-248006, India

<sup>2</sup>Department of Statistics, Kumaun University, SSS Campus, Almora, India-263601, India

## Article history

Received: 22-03-2017

Revised: 21-06-2017

Accepted: 19-07-2017

Corresponding author:

Girish Chandra

Division of Forestry Statistics,

Indian Council of Forestry

Research and Education,

Dehradun, India-248006, India

Email: gchandra23@yahoo.com

**Abstract:** An attempt has been made in this paper to improve the method ‘Square Random Lattice Sampling’ by suggesting a systematic procedure for achieving the goals under controls beyond stratification. The estimator of mean and its variance are calculated. Results shows that the relative precision of the proposed procedure increases as size of the latin square and the number of latin squares along the main diagonal of the population increases. This procedure seems to be a good alternative of the above method of for large population. Three different examples are also discussed to demonstrate the relative precision of the proposed method.

**Keywords:** Controlled Sampling, Latin Squares, Lattice Sampling, Relative Precision, Simple Random Sampling

## Introduction

Although the idea of lattice sampling was initially discussed by Patterson (1954; Yates, 1946; 1960), its need was long felt. The basic idea of lattice sampling is derived from ‘controls beyond stratification’. In many situations, it is desirable to stratify the population on the basis of more than one stratification variables (multi-stratification) which often leads to more strata cells than can be accommodated in a one-way stratified design. For an instance, when economic considerations restrict the number of primary selections to  $x$ , then  $x/2$  is the upper limit on the number of ordinary strata that can be created for paired selection, i.e., when variance estimation is required and  $x$  strata is the limit for single selection per stratum. However, sometimes it is required to have greater control over the selections than can be expressed with only  $x$  strata. For example, a national sample of interviews in the United States may need to be restricts to 100 counties, because of the cost of staffing counties. But to satisfy various requests for adequate controls over the distribution of sample counties would require many more strata than 100. These problems for national samples of counties are discussed by Frankel and Stock (1942; Goodman and Kish, 1950). A controlled selection of hospitals is described by Hess *et al.* (1961) whereas a controlled sample of cities of the U. S. Bureau of Labor Statistics (1961) for its price index by Stigler (1961).

Even a generation ago, conflicting needs of controls and randomization were widely thought to be unconflicting, as can be seen on the debate between

purposive and random methods in the Bulletin of the International Statistical Institute (1926) compiled by Jensen (1926). Demands for control often outrun the limitations of stratified random sampling, leading to the need for purposive selection, abandoning probability sampling all together. Several methods have been proposed by different authors for imposing more controls within the requirement of probability sampling. In fact, any departure from Simple Random Sampling (SRS) can be considered as a control. Again taking the example of stratified random sampling for the achieving control which provides us the opportunity to represents all the homogeneous subgroups within a heterogeneous population. Systematic sampling may be taken as another extreme of achieving controls in which only  $k$  samples in the sample space (preferred samples) having equal probabilities of selection  $1/k$  are left. The remaining  $({}^N C_n - k)$  samples (non-preferred samples) have zero probability of selection. In the similar way, two stage sampling is another method of achieving controls in which first stage units are selected with some specified probabilities, but the selection themselves are made on the basis of some prior restrictions (or controls).

Following the lead by Patterson (1954; Yates, 1946; 1960), the lattice sampling was further viewed by Jessen and co-authors under the label of “Lattice Sampling”, “Two-way Stratification”, “Deep Stratification” and “Multi-stratification”. In the next section those procedures are reviewed for drawing a sample, which permits cross classification restrictions to be met with less sample units than a one-way design.

## Controlled Sampling Procedures: Lattice Sampling

The basic idea for increasing the precision of the sampling design is to impose some restrictions (or controls) while selecting the sample. Due to these restrictions the probabilities of selection of preferred combination of units is increased and consequently decreases the selection probabilities of non-preferred (undesirable) combinations. In addition to the precision, three expected advantages of controlled sampling are shown by Hess and Srikanth (1966) as:

- Controls may be imposed to secure proper distribution geographically or otherwise and to ensure adequate sample size for subgroups that are domains of study
- To secure moderate reductions in the sampling errors of a multiplicity of characters simultaneously
- The significant reduction of the sampling error in the global estimates of specified key variable

Different selection procedures have been used by different authors. Some of the important works on this field are discussed as follows.

Frankel and Stock (1942) discussed the multi-stratification techniques in detail and studied their uses in data collection. In particular, they considered the possibility of using sample design in which the latin square principle can be used to reduce the number of sample units necessary to represent all strata. For example, suppose two criteria for stratification are used, say  $X$  and  $Y$ , such that  $L$  strata can be constructed from the  $X$  characteristic and, within each of these,  $L$  from the  $Y$  characteristic. If one relates the resulting pattern to a single treatment of an  $L \times L$  latin square, it is obvious that in a sample of  $L$  sample units of the  $LX$  strata will be represented and likewise each of the  $LY$  strata. Further, Tepping *et al.* (1943) discussed such designs under the title "Deep Stratification" in some detail and compared their variances with the variances of single stratification sampling.

Yates (1946) discussed the problem of selecting two-way stratified sample. Using ANOVA arguments, he suggested that the variance in two-way stratified sampling is smaller than that in the one-way stratified sampling. However, the scheme in the latter work of Yates (1953) resembles with the design scheme of Frankel and Stock (1942), in which the principle of latin square is used to select sample units. He also discussed the frame work in three-dimensional sampling scheme in which the additional dimension (strata) of vertical level (or file), can be represented in the sample by choosing  $r^2L$  units out of the  $L^3$  units in the population, if  $r$  units are replicated in each strata. Patterson (1954) extended

the work of Yates (1953), particularly with respect to the estimation of errors. He also suggested four methods for selecting samples with "Control" on both sets of strata in two-way stratification techniques. All these four methods are based on the number of replication from each row and each column and the type of lattices. However all four methods provide equal selection probabilities for each element in the population, therefore the sample mean will be an unbiased estimator of population mean for each method. But these four methods have distinguishing characteristics on sampling variance and in their ability to provide suitable estimates of sampling variances.

Goodman and Kish (1950) developed a procedure under the name "Controlled Selection" which assigns a probability of selection to each of several possible samples so that 'preferred' combinations of units are given a higher probability than 'non-preferred' combinations of units. Bryant *et al.* (1960) proposed a procedure under the title "Two-way Stratification" in which there are two stratifying criteria, both of which are desirable in a sample design. They used such procedure only when two variables are used for stratification and each has the same number of levels. In this situation, the number of permitted observations may be less than the number of strata formed by the usual double stratification technique. They also showed that if the stratification effects are additive in the analysis of variance sense, the method is particularly more useful. Bryant (1961) discusses various examples under "Multi-dimensional Controlled Selection" in which the number of strata cells exceeds the permissible sample size. In a survey of fish catch, he used four different types of strata viz. location, day time, season of summer and type of the day and likewise for others. He also extended the work of Bryant *et al.* (1960) for a well known pole mountain study (1961) in which he had developed the sampling and estimation technique. This study had four dimensions and the sample size  $n = 46$ . In this situation a four-dimensional cube of size  $46^4$  in population is required. The selection procedure is such that, the sample of cells is chosen in four dimensions so that once a row, column, level or flat is selected, it cannot be chosen again.

The techniques discussed above are based on the equal probability of selection of each strata cell. Jessen (1973) developed a technique under the name "Probability Lattice Sampling", in which each strata cell has unequal probability of selection, proportional to the size of the strata cells. Both equal probability and unequal probability methods were covered in Jessen's (1969; 1970; 1973; 1975) papers which are compiled by Jessen (1978). We therefore, restrict ourselves only to Jessen's work. First, we summarized the procedure of lattice sampling in the following subsections.

**Equal Probability Method**

When only one variable is used for stratification and each stratum (row and column) has equal number of levels, say  $L$ , there are  $L^2$  strata cells. In this situation, sample size  $n = rL$ , where  $r$  cells are to be selected from each row (column) of a square lattice of order  $L \times L$ . This procedure is called simple random lattice sampling. Suppose for a simple case, there are 16 elements arranged into 4 rows and 4 columns. We wish to draw a sample of size 4 (for  $r = 1$ ) from the  $4 \times 4$  universe. The particular lattice design, simple random simple stratification (using rows or column as strata), is given in Fig. 2.1.

An extension of simple random lattices is square random lattices. When two variables (row and column) are used for stratification and each has the same number of levels, say  $L$ , there are  $L^2$  cells in the design. Jessen (1975) proposed two techniques for drawing samples of size  $n = rL$ , where  $r$  cells are selected from each level of each variable (dimension). The first technique is ‘general lattice’ technique which starts with a selection of  $r$  cells at random from the  $L$  cells of the first row; likewise at random, select  $r$  cells from the  $L$  cells of the second row. This process is continues until  $r$  cells have been selected from one of the next row which contains only  $L-1$  permissible cells. The selection procedure continues until the last row has been covered such that from each column only  $r$  cells have been selected in the sample.

With this selection procedure, the variance of the sample mean per element,  $\bar{y}$  is given by:

$$Var(\bar{y})_{RL} = \frac{N-n}{Nn} S_{RC}^2 \tag{2.1}$$

where:

$$S_{RC}^2 = \frac{1}{(L-1)^2} \sum_{g=1}^L \sum_{h=1}^L E_{gh}^2$$

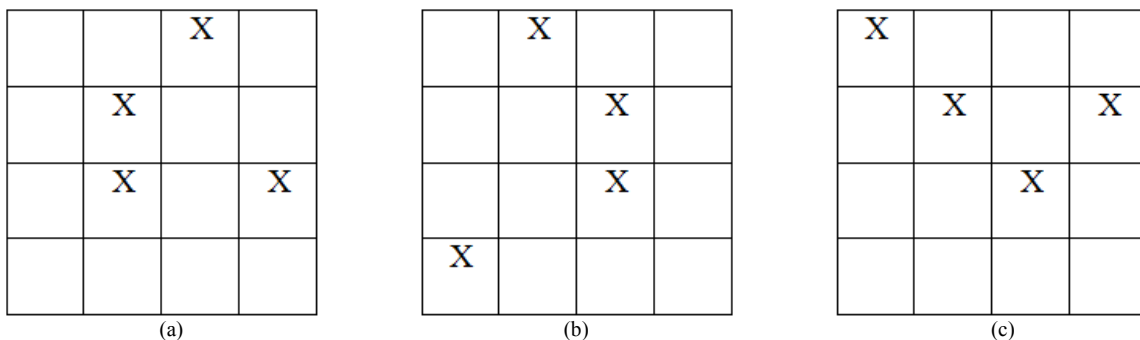


Fig. 1. Structure of different simple random lattice designs under (a) without stratification (simple random) (b) stratified row only (c) stratified column only

and where:

$$E_{gh} = y_{gh} - \bar{y}_g - \bar{y}_h + \bar{Y} \tag{2.2}$$

Where:

$y_{gh}$  = The value of the characteristic ( $y$ ) under study occupying row  $g$  and column  $h$ :

$\bar{y}_g = \frac{1}{L} \sum_{h=1}^L y_{gh}$  = The mean per cell of observations in row  $g$

$\bar{y}_h = \frac{1}{L} \sum_{g=1}^L y_{gh}$  = The mean per cell of observations in column  $h$

$\bar{Y} = \frac{1}{L^2} \sum_{g=1}^L \sum_{h=1}^L y_{gh}$  = The population mean

Generally, this scheme is a more efficient or very close to SRS or simple stratified (either row or column as stratification variable) sampling (Jessen, 1975).

The alternative technique, the ‘latin lattice’ is based on the assumption that  $L/r$  is a positive integer. Then  $r \times r$  squares are designated along the diagonal of the  $L \times L$  square, known as ‘latin squares’. In case of  $L$  is odd and  $r$  is even, the diagonal can be filled by using a combination of  $r \times r$  complete squares and  $(r+k) \times (r+k)$  incomplete squares for the appropriate value of  $k$ . This particular scheme (before randomization) may be improved by permuting rows and columns randomly to obtain the after randomization design. The selection scheme for these methods for a double lattice ( $r = 2$ ) is illustrated in the Fig. 2.2.

However, both the method gives identical results; we prefer latin lattice technique as an additional degree of freedom ( $r$  degrees of freedom) is available for the estimate of the variance (Jessen, 1975).

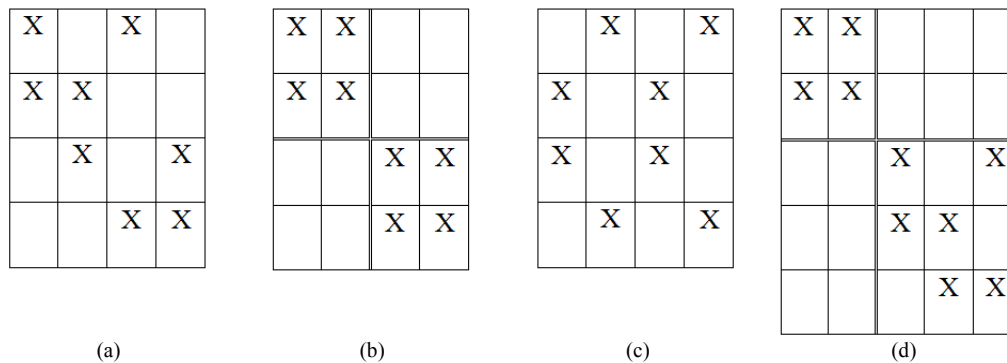


Fig. 2. Structure of lattices for  $r = 2$  under (a) general lattice of  $4 \times 4$  square (b) latin lattice of  $4 \times 4$  square before randomization of rows and columns (c) latin lattice of  $4 \times 4$  square after randomization (d) latin lattice of  $5 \times 5$  square before randomization ( $L/r$  is not a positive integer)

When two variables used for stratification have unequal number of levels, the square lattice concept can be extended to rectangular lattices of say  $R$  rows and  $C$  columns. Jessen (1978, chap. 11) discussed the method of selecting  $n = rt$ ,  $t$  being the larger of  $R$  and  $C$ , cells from such design. In such case; the latins generated along the diagonal will be rectangular or square, depending upon the dimensions of the rectangular population under consideration.

#### Unequal Probability Method

Sometimes the cells of a cross-classification contain an unequal number of units. It may then often desirable to sample the cells in a manner that reflects this unevenness. Such situations may arise where the sampling is done in two stages. First a sample of the cells is chosen and then a sample of units is drawn from the selected cells. Jessen (1970; 1973; 1975) discussed the situation where the cells of a two-way or a three-way cross-classification did not contain the same number of population units and it is desired to select the sample with the desired conditions which is termed as ‘unequal probability lattice’.

Each lattices made under this situations are often called ‘simple probability lattices’. The lattices can be extended to square probability lattices and cubic probability lattices by an additional dimension, although procedure becomes more complex by using three or more dimensions. If accuracy can be increased by extending stratification from one-way to two-way, then it appears that even further gains are possible by going on to three or even more dimensions, provided, that there are sensible factors to stratify with.

#### Marginal Stratification

In both the methods discussed above, it is assumed that the cell size is known in advance. When they are not, but the margin sizes are known to us to attempt some sort of stratification, taking account of the information on margin

sizes. This case arises when we may have given information on a series of one-way classification but none on two-way or crossed classification.

Tiwari and Nigam (1998) developed a procedure for selecting samples for random and probability lattices in square and rectangular frames without taking the restrictions  $n = rL$  and  $n = rt$  given by Jessen (1978) for square and rectangular lattices respectively. He also suggested estimates for the variance estimation for both the cases.

#### Systematic Square Random Lattice Sampling

When we deal either with the general lattice/latin lattice technique under square random lattice sampling, the following two limitations are encountered:

- There is no systematic procedure to achieves the sample from each row and each column, hence it is arbitrary, time consuming and costly.
- The method of general lattice requires to check the condition at every step after the procedure takes place in first  $(r-1)$  rows that  $r$  units (cells) being selected from each column. Also due to nature of the method of latin lattice (before randomization), some biasness may occurred. The method latin lattice (after randomization) is although preferable over the general lattice method, yet the number of steps involved in the procedure is quite large and there is no definite rule for permuting the rows and columns (after the case before randomization)

In the present paper, we have attempted these two issues observed in the Jessen’s work and suggested a convenient procedure for selecting samples using systematic sampling after applying the general lattice sampling with  $r = 1$  on the latin squares along the main diagonal. Following subsections deals with the proposed

method and comparison with SRS and square random lattice sampling.

*The Proposed Plan*

As we know the method of ‘latin lattice’ is preferred over the ‘general lattice’ due to the availability of an additional degree of freedom for latin lattice while estimating the variance. In this section, we have proposed an improved latin lattice method which overcomes two limitations, as discussed above.

Let us consider a two-way population frame consisting of  $N = L \times L$  units. A sample of size  $n$  is to be drawn from this population utilizing two stratification variables (row and column) and satisfy the two basic restrictions, viz:

- $N = r \times L$ ,  $r$  being the number of units (cells) replicates in each row and each column
- $\frac{N}{n} = k$ ,  $k$  being a positive integer

Here, the whole population may be partitioned into  $r$  equal parts on the basis of rows and columns both. This results  $r^2$  latin squares of same size  $k \times k$  which are shown in matrix form (Equation 3.1) as:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} & \dots & a_{1L} \\ a_{21} & a_{22} & \dots & a_{2j} & \dots & a_{2L} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{i1} & a_{i2} & \dots & a_{ij} & \dots & a_{iL} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{L1} & a_{L2} & \dots & a_{Lj} & \dots & a_{LL} \end{bmatrix} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1i} & \dots & A_{1r} \\ A_{21} & A_{22} & \dots & A_{2i} & \dots & A_{2r} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{i1} & A_{i2} & \dots & A_{ii} & \dots & A_{ir} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{r1} & A_{r2} & \dots & A_{ri} & \dots & A_{rr} \end{bmatrix} \quad (3.1)$$

where,  $A_{ij}$  is the sub-matrix of  $a_{ij}$ 's whose first element is  $a_{(i-1)k+1, (j-1)k+1}$ .

Now, the proposed selection procedure of units, consisting of two steps, is described as follows:

**Step 1:** Select  $k$  units from each latin squares placed along the main diagonal of  $A$ , i.e.,  $A_{11}, A_{22}, \dots, A_{rr}$  using the general lattice technique (i.e.,  $r = 1$ ).

**Step 2:** Select every  $k$ th unit using systematic sampling from the selected units in the Step 1 along the corresponding rows and columns of the whole population frame. This step shall cover all the latin squares in addition to the latin square along the main diagonal.

Under this procedure, there are a total of  $[2(r-1)+1]k$ ,  $[2(r-2)+1]k$ ,  $[2(r-3)+1]k, \dots, k$  units selected in the sample utilizing latin squares  $A_{11}, A_{22}, A_{33}, \dots, A_{rr}$  respectively. Accordingly, total sample size  $n$  in terms of  $r$  and  $k$  is given by:

$$n = [2(r-1)+1]k + [2(r-2)+1]k + \dots + k = r^2k$$

Consequently, population size  $N = r^2k^2$ .

Further, the probabilities of selecting the units from the rows/columns of latin squares  $A_{11}, A_{22}, A_{33}, \dots, A_{rr}$  will be  $1/k, 1/(k-1), 1/(k-2), \dots, 1/2, 1$  respectively.

To illustrate this procedure, let us consider a  $6 \times 6$  population  $A = (a_{ij})_{6 \times 6}$  with  $r = 2$  and  $k = 3$ . This results total 4 possible latin squares of size  $3 \times 3$ . In the first step of this example, the units  $a_{13}, a_{21}, a_{32}, a_{44}, a_{56}$  and  $a_{65}$  are selected from the latin squares  $A_{11}$  and  $A_{22}$  using the method of general lattices. The step 2 starts with selecting every 3rd ( $k = 3$ ) unit systematically along each row and each column of matrix  $A$  from the selected units under step 1. The units selected under step 1 and step 2 are shown in Fig. 3. In this particular example, the step 2 cannot be applicable for  $A_{22}$  due to non-availability of cells for selection.

The total number of possible samples in this proposed procedure will depend on arrangement of cells of latin squares  $A_{11}, A_{22}, \dots, A_{rr}$ . For  $r = 1$ , there are a total of  $k!$  possible combination of cells to be selected from each of these latin squares under step 1. Hence, in general, total number of possible samples shall be  $(k!)^r$ .

*Mean and Variances*

For the sake of convenience in mathematical calculations, the matrix  $A$  (Equation 3.1) can be rewritten as Equation 3.2, in which latin squares are numbered continuously according to the rows only. This is done by substituting 1 in place of first subscript and  $\{r(\text{first subscript} - 1) + \text{next subscript}\}$  in place of next subscript:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1i} & \dots & A_{1r} \\ A_{1(r+1)} & A_{1(r+2)} & \dots & A_{1(r+3)} & \dots & A_{1(2r)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{1\{(i-1)r+1\}} & A_{1\{(i-1)r+2\}} & \dots & A_{1\{(i-1)r+i\}} & \dots & A_{1(ir)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ A_{1\{(r-1)r+1\}} & A_{1\{(r-1)r+2\}} & \dots & A_{1\{(r-1)r+i\}} & \dots & A_{1(r^2)} \end{bmatrix} \quad (3.2)$$

Now, we use the following notations:

- $y_{ij}$  = The value of the characteristic under study of  $j$ th unit of the  $i$ th latin square ( $i = 1, 2, \dots, r^2, j = 1, 2, \dots, k^2$ )
- $\bar{Y}_i$  = The population mean of the  $i$ th latin square =  $\frac{1}{k^2} \sum_{j=1}^{k^2} y_{ij}$
- $\bar{y}_i$  = The sample mean of  $i$ th latin square =  $\frac{1}{k} \sum_{j=1}^k y_{ij}$
- $\bar{Y}$  = Overall population mean per element =  $\frac{1}{r^2} \sum_{i=1}^{r^2} \bar{Y}_i$
- $\bar{y}$  = Sample mean per element =  $\frac{1}{r^2} \sum_{i=1}^{r^2} \bar{y}_i$

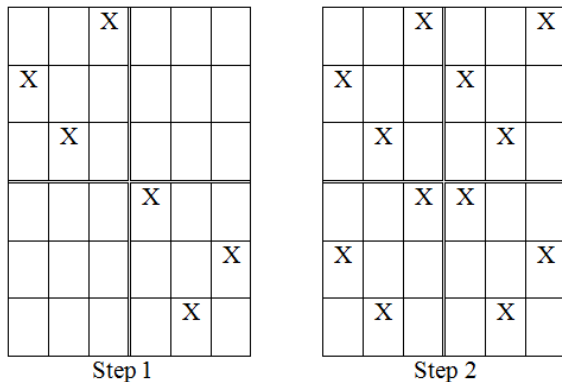


Fig. 3. Selected Units under the proposed plan for the 6x6 population with  $r = 2$  and  $k = 3$

To show  $\bar{y}$  as an unbiased estimator of  $\bar{Y}$ , it is to show that  $\bar{y}_i$  is an unbiased estimator of  $\bar{Y}_i$ . For which we have:

$$E(\bar{y}_i) = \frac{1}{k} \sum_{j=1}^k E(y_{ij}) = \frac{1}{k} \sum_{j=1}^k \frac{1}{k^2} \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} y_{ij} = \bar{Y}_i$$

Therefore:

$$E(\bar{y}) = \frac{1}{r^2} \sum_{i=1}^{r^2} E(\bar{y}_i) = \bar{Y}$$

Now, all the latin squares under the proposed procedure are random and may be taken as a single unit and therefore the variance of  $\bar{y}$  is calculated in parallel with SRS and is written as:

$$Var(\bar{y})_{SL} = \frac{N-n}{Nn} S_{BL}^2 \tag{3.3}$$

where, subscript ‘SL’ represents for ‘Systematic Lattice’ and  $S_{BL}^2$  is the variance between latin squares as:

$$S_{BL}^2 = \frac{1}{r^2-1} \sum_{i=1}^{r^2} (\bar{Y}_i - \bar{Y})^2 \tag{3.4}$$

The unbiased estimator of  $S_{BL}^2$  is given by:

$$\hat{S}_{BL}^2 = s_{bl}^2 = \frac{1}{r^2-1} \sum_{i=1}^{r^2} (\bar{y}_i - \bar{y})^2$$

**Comparison with SRS**

The variance of sample mean under SRS is given by:

$$Var(\bar{y})_{SRS} = \frac{N-n}{Nn} S^2$$

where:

$$S^2 = \frac{1}{r^2 k^2 - 1} \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} (y_{ij} - \bar{Y})^2$$

Or:

$$\begin{aligned} (r^2 k^2 - 1) S^2 &= \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} [(y_{ij} - \bar{Y}_i) + (\bar{Y}_i - \bar{Y})]^2 \\ &= \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} (y_{ij} - \bar{Y}_i)^2 + \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} (\bar{Y}_i - \bar{Y})^2 \end{aligned} \text{ (Since cross product vanishes)}$$

Or:

$$(r^2 k^2 - 1) S^2 = (k^2 - 1) \sum_{i=1}^{r^2} S_i^2 + k^2 (r^2 - 1) S_{BL}^2 \tag{3.5}$$

where,  $S_i^2 = \frac{1}{k^2} \sum_{j=1}^{k^2} (y_{ij} - \bar{Y}_i)^2$  represents the within latin square variance of the  $i$ th latin square.

From the Equation 3.5) it is clear that for large values of  $r$  and  $k$ :

$$\begin{aligned} S^2 &= \frac{1}{r^2} \sum_{i=1}^{r^2} S_i^2 + S_{BL}^2 \\ \Rightarrow S^2 &> S_{BL}^2 \end{aligned} \tag{3.6}$$

This implies that the proposed procedure is always more efficient than SRS for large values of  $r$  and  $k$ . Sometimes, for small values of  $r$  and  $k$  the proposed procedure may be weaker than the SRS in terms of relative precision (as shows in the Example 1 in the next section). The Equation 3.6 also indicates that as within latin square variance increases and (or) between latin square variance decrease, the Relative Precision (RP) with respect to SRS increases rapidly.

**Comparison with Square Random Lattice Sampling**

The variance of  $\bar{y}$  under square random lattice sampling is given in Equation 2.1 in which we have:

$$S_{RC}^2 = \frac{1}{(L-1)^2} \sum_{g=1}^L \sum_{h=1}^L E_{gh}^2 \tag{3.7}$$

In this section, our aim is to compare  $S_{RC}^2$  with  $S_{BL}^2$ .

On simplifying Equation 3.7 and using the relation  $\bar{Y}_g = \frac{1}{L} \sum_{h=1}^L y_{gh}$ ,  $\bar{Y}_h = \frac{1}{L} \sum_{g=1}^L y_{gh}$  and  $\bar{Y} = \frac{1}{L} \sum_{g=1}^L \bar{Y}_g = \frac{1}{L} \sum_{h=1}^L \bar{Y}_h$  we get:

$$S_{RC}^2 = \frac{1}{(L-1)^2} \left[ \sum_{g=1}^L \sum_{h=1}^L y_{gh}^2 - L \sum_{g=1}^L \bar{Y}_g^2 - L \sum_{h=1}^L \bar{Y}_h^2 + L^2 \bar{Y}^2 \right] \quad (3.8)$$

Similarly, on simplifying Equation 3.4 we get:

$$S_{BL}^2 = \frac{1}{k^4(r^2-1)} \left[ \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} y_{ij}^2 + 2 \sum_{i=1}^{r^2} \sum_{j < j'=1}^{k^2} \sum_{j'=1}^{k^2} y_{ij} y_{ij'} \right] - \frac{r^2}{r^2-1} \bar{Y}^2 \quad (3.9)$$

We know that  $L = rk$ .

For  $k \geq r$  it is clear that the first term of (3.7) is greater than the first term of (3.8) as:

$$\frac{1}{(L-1)^2} > \frac{1}{k^4(r^2-1)}$$

and:

$$\sum_{g=1}^L \sum_{h=1}^L y_{gh}^2 = \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} y_{ij}^2$$

Therefore, the  $S_{RC}^2 > S_{BL}^2$  is always hold if the population satisfies the following inequality:

$$\begin{aligned} & \frac{(L)^2}{(L-1)^2} \left[ \bar{Y}^2 - \frac{1}{L} \sum_{g=1}^L \bar{Y}_g^2 - \frac{1}{L} \sum_{h=1}^L \bar{Y}_h^2 \right] \\ & > \frac{1}{(r^2-1)} \left[ \frac{2}{k^4} \sum_{i=1}^{r^2} \sum_{j < j'=1}^{k^2} \sum_{j'=1}^{k^2} y_{ij} y_{ij'} - r^2 \bar{Y}^2 \right] \end{aligned} \quad (3.10)$$

This comparison shows that the RP with respect to square random lattice sampling increases as the value of  $k$  increases more than  $r$ .

### Numerical Examples

In this section, we have considered three examples (4x4, 6x6 and 9x9) to show the performance of the proposed systematic square random lattice sampling in comparison with SRS and square random lattice sampling.

#### Example 1

This small example with  $r = k = 2$  is borrowed from Jessen (1975). The values of characteristic under study (y-values) are as:

-5	-6	-4	-1
-3	-2	-2	-1
-1	2	4	3
1	2	6	7

In this population having 16 units, we wish to select 8 units by all the three procedures under study. The 4

possible latin squares of size 2x2 each are shown in the following Table 1.

For variance calculation, we have:

$$\begin{aligned} \sum_{g=1}^L \sum_{h=1}^L y_{gh}^2 &= 216, \sum_{g=1}^L \bar{Y}_g^2 + \sum_{h=1}^L \bar{Y}_h^2 = 50, \bar{Y} = 0, \\ \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} y_{ij}^2 + 2 \sum_{i=1}^{r^2} \sum_{j < j'=1}^{k^2} \sum_{j'=1}^{k^2} y_{ij} y_{ij'} &= 256 \end{aligned}$$

and:

$$\bar{Y}_1 = -4, \bar{Y}_2 = -2, \bar{Y}_3 = 1, \bar{Y}_4 = 5$$

Therefore:

$$S^2 = 14.40, S_{RC}^2 = 1.7778 \text{ and } S_{BL}^2 = 15.33$$

Which results:

$$Var(\bar{y})_{RL} = 0.1111, Var(\bar{y})_{SRS} = 0.90 \text{ and } Var(\bar{y})_{SL} = 0.9583$$

The values of these variances shows that for small values of  $r$  and  $k$  the proposed procedure is not performing well in comparison to SRS and square random lattice sampling. This performance can be calculated in terms of RP as:

RP with respect to square random lattice sampling is:

$$RP_{RL} = \frac{Var(\bar{y})_{RL}}{Var(\bar{y})_{SL}} = 0.1159$$

Similarly, RP with respect to SRS is:

$$RP_{SRS} = 0.9391$$

#### Example 2

This example consist of 6x6 frame simulated from the standard normal distribution for selecting the sample of size 12 with  $r = 2, k = 3$  using SPSS software. The data is shown in the following Table 2.

For this example, we get:

$$\begin{aligned} \sum_{g=1}^L \sum_{h=1}^L y_{gh}^2 &= 16.226, \sum_{g=1}^L \bar{Y}_g^2 + \sum_{h=1}^L \bar{Y}_h^2 = 0.8146, \bar{Y} = -0.0768 \\ \sum_{i=1}^{r^2} \sum_{j=1}^{k^2} y_{ij}^2 + 2 \sum_{i=1}^{r^2} \sum_{j < j'=1}^{k^2} \sum_{j'=1}^{k^2} y_{ij} y_{ij'} &= 2.4341 \text{ and} \\ \bar{Y}_1 &= -0.0407, \bar{Y}_2 = -0.1255, \bar{Y}_3 = -0.1072, \bar{Y}_4 = -0.0339 \end{aligned}$$

These calculations results us:

$$S^2 = 0.4575, S_{RC}^2 = 0.4621 \text{ and } S_{BL}^2 = 0.0022$$

Table 1. Layout of 4 possible latin squares of size 2x2 each

Rows (g)	Columns (h)				Row mean	
	1	2	3	4	$\bar{Y}_g$	$\bar{Y}_g^2$
1	-5	-6	-4	-1	-4	16
2	-3	-2	-2	-1	-2	4
3	-1	2	4	3	2	4
4	1	2	6	7	4	16
Column mean						
$\bar{Y}_h$	-2	-1	1	2	0	
$\bar{Y}_h^2$	4	1	1	4		

Table 2. The y-values of the 6x6 from standard normal distribution

-1.3652	-0.8934	1.2334	0.9012	0.7877	0.1273
0.5918	0.2370	0.4116	0.3310	-0.5938	-0.2653
-0.1800	-0.3364	-0.0654	-0.2369	-1.6175	-0.5633
-1.2970	-0.1548	0.4089	-1.2669	0.4677	0.4613
-0.4345	-0.3022	0.7202	-0.2451	-0.5500	0.7937
0.2890	0.2175	-0.4117	-0.1795	-0.2738	0.4873

Table 3. y-values of the 9x9 population generated from Uniform (0, 10) distribution

1.7109	4.3388	2.7824	2.3011	3.8157	8.8104	2.8916	0.5832	5.2263
7.8249	5.6612	2.8727	5.6227	4.8314	1.3929	6.0027	1.1777	2.3252
8.7848	7.2509	4.1758	7.6672	4.1493	1.7008	2.7927	7.7853	3.1529
6.4714	2.7421	2.7265	5.0514	2.1833	7.2094	7.7212	0.3833	0.7367
1.2980	9.7015	3.1468	8.7091	5.8736	4.5854	1.0385	7.5481	3.4635
9.4034	0.8823	8.3145	6.2188	9.1574	6.1586	9.7812	9.5111	6.8108
3.5978	9.0875	4.5961	5.5031	8.0544	9.1189	3.4599	3.2472	3.2942
1.6904	0.4532	2.1876	5.4253	4.6825	5.8467	9.7302	8.1790	5.9426
1.7646	9.7351	9.7259	7.6794	1.1438	7.8930	2.9209	4.9190	1.6779

Which implies that:

$$Var(\bar{y})_{SL} = 0.00012, Var(\bar{y})_{SRS} = 0.02542$$

$$and Var(\bar{y})_{RL} = 0.02567$$

Therefore, the RPs are calculated as:

$$RP_{RL} = 215.41, RP_{SRS} = 213.31$$

### Example 3

The higher order population of size 9x9 is generated from the uniform distribution U(0, 10) as shown in Table 3. For this example, the sample size is taken 27 with  $r = k = 3$ .

The final calculation of variances is comes out to be:

$$Var(\bar{y})_{SL} = 0.0158, Var(\bar{y})_{SRS} = 0.2039 and$$

$$Var(\bar{y})_{RL} = 0.2057$$

and therefore:

$$RP_{RL} = 13.03, RP_{SRS} = 12.92$$

### Conclusion

We propose a method of selecting samples under the two way stratification structure of the population. In many forest surveys, it is observed that the forest area (generally taken as a forest stand which may be national park, biosphere reserve etc having area more than 500 sq. km) under study is very large. The sampling is usually done by quadrat (of size say 0.1 hectare) method under two way stratification. In such situation, the number of rows and columns are increases instantly as forest area increases. Therefore, a systematic procedure is required without checking the number of units selects from each row and column in every steps. The present procedure may be helpful under such situation when population size is large. In this procedure, we require only the latin square placed across the main diagonal of the population for applying the systematic sampling for selecting the units from other latin squares. The second advantage of the large population size is that the increase of relative precision of the proposed procedure as population size increases. We recommend that the size of the latin square should be larger than the number of



latin squares placed across the main diagonal of the population. Finally, it may be said that the proposed procedure is a good alternative of the procedure of squared random lattice sampling as proposed by Jessen (1975) for large population.

## Acknowledgment

The authors thank the unknown Reviewers and the Editor for their useful comments and suggestions in the earlier version of this article.

## Author's Contributions

**Girish Chandra:** Derived the scheme and formulae for unbiased estimators and their variances for the mean under square random lattice sampling and did all computations of relative precisions of the estimators.

**Neeraj Tiwari:** Suggested to improve the difficulty on square random lattice sampling for the large population size and reviewed the literature on lattice sampling. He did overall drafting and editing of the paper.

## Ethics

The authors confirm that they have read and approved the manuscript and declare that no ethical issues are involved in it.

## References

- Bryant, E.C., 1961. Sampling methods. Seminar Papers, Iowa State University.
- Bryant, E.C., H.G. Hartley and R.J. Jessen, 1960. Design and estimation in two-way stratification. *J. Am. Stat. Assoc.*, 55: 105-124.  
DOI: 10.1080/01621459.1960.10482052
- Frankel, L.R. and J.S. Stock, 1942. On the sample survey of unemployment. *J. Am. Stat. Assoc.*, 10: 288-293.  
DOI: 10.1080/01621459.1942.10500615
- Goodman, R. and L. Kish, 1950. Controlled selection-a technique in probability sampling. *J. Am. Stat. Assoc.*, 45: 350-372.  
DOI: 10.1080/01621459.1950.10501130
- Hess, I., D.C. Riedel and T.B. Fitzpatrick, 1961. Probability sampling of hospitals and patients. *Ann. Arbor*, Mich.: Bureau of Hospitals Administration: The University of Michigan, Bureau of Hospital Administration.
- Hess, J. and K.S. Srikanthan, 1966. Some aspects of the probability sampling technique of controlled selection. *Health Serv. Res.*, 1: 8-52.
- Jensen, A., 1926. Report on the representative method in statistics. *Bull. Int. Stat. Inst.*, 22: 359-380.
- Jessen, R.J., 1969. Some methods of probability non-replacement sampling. *J. Am. Stat. Assoc.*, 64: 175-193. DOI: 10.1080/01621459.1969.10500962
- Jessen, R.I., 1970. Probability sampling with marginal constraints. *J. Am. Stat. Assoc.*, 65: 776-796.  
DOI: 10.1080/01621459.1970.10481124
- Jessen, R.J., 1973. Some properties of probability lattice sampling. *J. Am. Stat. Assoc.*, 68: 26-28.  
DOI: 10.1080/01621459.1973.10481328
- Jessen, R.J., 1975. Square and cubic lattice sampling. *Biometrics*, 31: 449-471. DOI: 10.2307/2529431
- Jessen, R.J., 1978. *Statistical Survey Techniques*. 1st Edn., John Wiley and Sons, New York, ISBN-10: 0471442607, pp: 520.
- Patterson, H.D., 1954. The errors of lattice sampling. *J. Royal Stat. Society*, 16: 140-149.
- Stigler, G., 1961. The price statistics of the federal government. Report to the Office of Statistical Standards, Bureau of the Budget, National Bureau of Economic Research, New York.
- Tepping, B.J., W.N. Hurwitz and W.E. Deming, 1943. On the efficiency of deep stratification in block sampling. *J. Am. Stat. Assoc.*, 45: 93-100.  
DOI: 10.1080/01621459.1943.10501781
- Tiwari, N. and A.K. Nigam, 1998. On two-dimensional optimal controlled selection. *J. Stat. Plann. Inference*, 69: 89-100.  
DOI: 10.1016/S0378-3758(97)00128-6
- Yates, F., 1946. A review of recent statistical developments in sampling and sampling surveys. *J. Royal Stat. Society*, 109: 12-30.  
DOI: 10.2307/2981390
- Yates, F., 1953. *Sampling Methods for Censuses and Surveys*. 2nd Edn., Carles Griffin and Company, London, pp: 401
- Yates, F., 1960. *Sampling Methods for Censuses and Surveys*. 3rd Edn., Carles Griffin and Company, London, pp: 440.