J. Ind. Soc. Agril. Statist.
60(2), 2006 : 85–89

Lattice Sampling without Sample Size Restriction

Neeraj Tiwari
SDPG College, Muzaffarnagar
(Received : August, 1993)

SUMMARY

A problem quite similar to controlled selection problem, was studied by Jessen in a series of papers under the label of 'Lattice Sampling'. His methods have severe sample size restriction, are ad-hoc, involving trial and error and may even fail to generate the set of feasible samples for probability lattices. Further, in the case of probability lattices, the unbiased and non-negative estimation of variance is not possible in many general situations, if as per Jessen's recommendations, the Horvitz-Thompson and Yates-Grundy form of variance estimators are used.

We advance here a unified method which does not have any sample size restriction and it also facilitates us the non-negative estimation of variance. The method derives its strength from the idea of 'latin lattices' of Jessen (1975).

Key words : Controlled selection, Controlled rounding, Random lattices, Probability lattices, Cross-stratification.

1. INTRODUCTION

In many situations, it is desirable to stratify the population on the basis of more than one stratification variables. Such a multiple stratification often leads to more strata cells than can be accommodated in a one-way stratified design. Goodman and Kish (1950), Bryant *et al.* (1960), Hess and Srikantan (1966) and Jessen (1970, 1973, 1975, 1978) have proposed various procedures for drawing the sample that permits cross-stratification restriction to be satisfied with less sample units than in a traditional one-way stratified design. The same problem has been addressed by Ernst (1981), Causey *et al.* (1985) and Cox (1987), wherein, they have attempted the solution to the problem through controlled rounding and transportation theory. This approach only gives a solution to obtaining the required sample and does not reveal anything about the related estimation problems.

We concern ourselves with the approach of Jessen. One of the main limitations of Jessen's approach is that for square and cubic lattices of order L, the sample size (n) must be r.L and r^2.L, respectively, where r is the number of strata cells selected from each row and column. Similarly, for rectangular lattices of order R×C, with R rows and C columns, the sample size must be r.t, where t is R or C, whichever is larger.

In many situations, specially, where the sampling is being done in two stages, the cells may contain an unequal number of units and it is desirable to sample the cells in a way that the unevenness of cells is given due consideration. The lattices obtained from the sample units selected in such situations are known as 'probability lattices'. For such cases, Jessen (1970) discussed two methods and referred them as 'Method 2' and 'Method 3'. These methods are restricted to a sample of size n = r.L. Similar methods were again discussed by Jessen (1975, 1978). Jessen (1978) modified his methods 2 and 3 to suggest an adaptive method, which he recommended for use in practice. However, the method is quite arbitrary and may fail to generate a set of feasible samples even for the case when n = r. L, as illustrated at the end of Section 3 of this article. Further, the method runs into difficulties for estimation of the variance. Jessen recommends the use of Horvitz-Thompson and Yates-Grundy form of variance estimators. However, his adaptive method fails to ensure the conditions $\pi_{ij} > 0$ and $\pi_{ij} \le \pi_i \pi_j$, required for unbiased and non-negative estimation of variance, π_i and π_{ij} being inclusion probabilities of first and second order, respectively.

We advance here a unified method which is also applicable when $n \ne r.L$ and $n \ne r.t$ for square and rectangular lattice frameworks, respectively. The method derives its strength from the idea of 'latin lattice', originally proposed by Jessen for equal probability lattices. The proposed method is superior to earlier methods of Jessen, as it facilitates estimation of variance more accurately. To gain insight, we first deal with equal probability lattices.

2. SIMPLE RANDOM LATTICES OF N = L² WITH n ≠ r.L

For the class of equal probabilities, various randomization schemes may be used and the lattices thus obtained from the sample units selected are termed as 'random lattices'. We consider the situation where two variables are used for stratification and each has the same number of levels, say, L. We wish to select a sample of size n, without the restriction that n = r. L.

For square equal probability lattices, Jessen advanced two methods for selecting a sample of size n = r. L. These are 'general lattices' and 'latin lattices'. Out of the two, the latin lattices are preferable as an additional degree of freedom is available for the estimate of the variance. Jessen's method of latin lattices is to take L/r, if L/r is an integer, r × r latins along the diagonal. When $n \ne r.L$, this can be modified as follows.

We take m latin squares of order L/m, with L/m an integer, along the diagonal, with some latin squares being either incomplete like Youden squares or with missing values. Even if L/m is not an integer, the same procedure can be followed with one latin square, like in Jessen's approach, being of higher dimension.

To explain the procedure, consider the case of L = 6 and n = 8. Two alternative arrangements with m = 3 are displayed below.

Arrangement 1

A	B	×	×	×	×
B	A	×	×	×	×
×	×	×	×	×	×
×	×	×	×	×	×
×	×	×	×	A	×
×	×	×	×	×	A

Arrangement 2

A	B	×	×	×	×
×	A	×	×	×	×
×	×	A	B	×	×
×	×	×	A	×	×
×	×	×	×	A	×
×	×	×	×	×	A

It can easily be seen that the above lattices retain all the properties of latin lattices of Jessen (1975).

Let $\bar{\bar{y}}_q$ be the mean of the q-th (q = 1, ..., m) latin, then the overall sample mean will be given by

$$\bar{\bar{y}} = (1/m)\sum_{q=1}^{m} \bar{\bar{y}}_q$$

and its variance will be given by

$$\text{Var}(\bar{\bar{y}}) = (1-f)S_{RC}^2/n$$

where f = n/N and

$$S_{RC}^2 = \sum_{i}^{L} \sum_{j}^{L} (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y})^2/(L-1)^2$$

Y_{ij} = the observed value of y for the element at the i^th^ row and j^th^ column

$\bar{Y}_{i.}$ = the mean per element of observations in the i^th^ row $= \sum_{j=1}^{L} Y_{ij}/L$

$\bar{Y}_{.j}$ = the mean per element of observations in the j^th^ column $= \sum_{i=1}^{L} Y_{ij}/L$

$\bar{Y}$ = the overall population mean $= \sum_{i}^{L} \sum_{j}^{L} Y_{ij}/L^2$

Along the lines suggested in Jessen (1975), it can be established that the mean square within latins for sub-classes provides an unbiased estimate for the interaction S_{RC}^2. We can form the following table for, say, Arrangement 1:

		Replicate		
		A	B	Total
	1	Y_{1A}	Y_{1B}	Y_1
Latin	2	Y_{2A}	–	Y_2
	3	Y_{3A}	–	Y_3

Here Y_{ix} is the total for the X[th] letter in the i[th] latin.

The estimate $\hat{S}_{RC}^2$ is given by

$$\hat{S}_{RC}^2 = \left[\frac{Y_{1A}^2}{2} + \frac{Y_{1B}^2}{2} + \frac{Y_{2A}^2}{2} + \frac{Y_{3A}^2}{2} - \left(\frac{Y_1^2}{4} + \frac{Y_2^2}{2} + \frac{Y_3^2}{2}\right)\right] \Big/ 3$$

Similarly, for Arrangement 2, the estimate $\hat{S}_{RC}^2$ is given by

$$\hat{S}_{RC}^2 = \left[\frac{Y_{1A}^2}{2} + Y_{1B}^2 + \frac{Y_{2A}^2}{2} + Y_{2B}^2 + \frac{Y_{3A}^2}{2} - \left(\frac{Y_1^2}{3} + \frac{Y_2^2}{3} + \frac{Y_3^2}{2}\right)\right] \Big/ 3$$

The estimate of the variance of the sample mean per element ($\bar{\bar{y}}$) can now easily be obtained by using

$$\hat{V}ar(\bar{\bar{y}}) = (1-f)\frac{\hat{S}_{RC}^2}{n}, \text{ where } f = n/N \qquad (2.1)$$

3. PROBABILITY LATTICES OF N= L² WITH n ≠ r. L

Let M_{ij} be the number of elements in the ij-th strata cell (i = j = 1, ..., L) and M be the total number of elements in the universe and $A_{ij} = M_{ij}/M$. We obtain the expected number of elements, n. Aij, for each of the ij-th cell and the row, column and grand totals. Now we select a feasible sample that meets the cell and marginal requirements, preferring the cells having largest n. A_{ij}'s to be selected first. While selecting the feasible samples, we also take care that, before or after randomization, the selected sample can be put into the form of m latin squares (either complete or incomplete with missing values) along the diagonal, as discussed in Section 2. This extra caution will help us in obtaining the estimate of the variance. Designate the selected cells with the asterisk (*). If n.A_{ij}'s for the problem under consideration are of the order 1/10, then 0.1 is subtracted from each of the designated n. A_{ij}'s and the process is repeated until the last feasible sample, meeting the cell and marginal requirements, is obtained. The process terminates after 10 feasible samples, each with selection probability 0.1, are obtained. However, if the n. A_{ij}'s are of the order 1/100, then instead of 0.1, 0.05 is subtracted from the designated n. A_{ij}'s and the process terminates after 20 feasible samples, each with selection probability 0.05, are obtained.

We borrow a two-way frame from Jessen (1978, 11.8) [Case A, Fig. 11.3, p. 373] to explain the method of selecting a sample of size n, where n ≠ r. L. The two-way frame is in the form of a 4 × 4 square having equal integer margins, with the following M_{ij}'s and totals.

0	3	5	2	10
4	2	2	2	10
3	1	2	4	10
3	4	1	2	10
10	10	10	10	40

We wish to select a sample of size 6 from this universe. The n. A_{ij}'s and their totals will be

0	0.45	0.75	0.30	1.5
0.60	0.30	0.30	0.30	1.5
0.45	0.15	0.30	0.60	1.5
0.45	0.60	0.15	0.30	1.5
1.5	1.5	1.5	1.5	6.0

Here, we have to select at least 1 and at the most 2 cells from each row and column so that a total of 6 elements are chosen at each feasible sample. We subtract 0.05 from each designated n. A_{ij}'s and a total of 20 feasible samples (some of which are duplicates, that is, containing the same elements), each of size 6, are obtained, which are demonstrated below.

(1)

0	.45*	.75*	.30
.60*	.30	.30	.30
.45*	.15	.30	.60*
.45	.60	.15	.30

(2)

0	.40*	.70*	.30
.55*	.30	.30	.30
.40*	.15	.30	.55*
.45	.55*	.15	.30

(3)

0	.35*	.65*	.30
.50*	.30	.30	.30
.35*	.15	.30	.50*
.45	.50*	.15	.30

(4)

0	.30*	.60*	.30
.45*	.30	.30	.30
.30*	.15	.30	.45*
.45	.45*	.15	.30

(5)

0	.25*	.55*	.30
.40*	.30	.30	.30*
.25	.15	.30	.40*
.45	.40*	.15	.30

(6)

0	.20	.50*	.30
.35	.30*	.30	.25
.25*	.15	.30	.35*
.45*	.35	.15	.30*

(7)

0	.20	.45*	.30
.35*	.25	.30	.25
.20	.15	.30*	.30*
.40*	.35*	.15	.25

(8)

0	.20	.40*	.30
.30*	.25	.30	.25
.20	.15	.25*	.25*
.35*	.30*	.15	.25

(9)

0	.20	.35*	.30*
.25	.25	.30*	.25
.20*	.15	.20	.20
.30*	.25*	.15	.25

(10)

0	.20	.30*	.25
.25	.25*	.25*	.25
.15	.15	.20	.20*
.25*	.20	.15	.25*

(11)

0	.20	.25*	.25*
.25	.20	.20	.25*
.15	.15*	.20	.15
.20*	.20*	.15	.20

(12)

0	.20*	.20*	.20
.25*	.20	.20	.20*
.15*	.10	.20	.15
.15	.15*	.15	.20

(13)

0	.15	.15	.20*
.20	.20*	.20*	.15
.10	.10	.20*	.15
.15*	.10	.15	.20*

(14)

0	.15*	.15*	.15
.20*	.15	.15	.15*
.10	.10	.15	.15*
.10	.10	.15*	.15

(15)

0	.10*	.10	.15
.15*	.15*	.15	.10
.10	.10	.15*	.10
.10	.10	.10*	.15*

(16)

0	.05	.10	.15*
.10	.10	.15*	.10*
.10*	.10*	.10	.10
.10	.10*	.05	.10

(17)

0	.05	.10*	.10
.10	.10*	.10*	.05
.05	.05	.10	.10*
.10*	.05	.05	.10*

(18)

0	.05	.05	.10*
.10*	.05	.05	.05*
.05	.05*	.10*	.05
.05	.05	.05*	.05

(19)

0	.05	.05*	.05*
.05	.05*	.05	–
.05	–	.05	.05*
.05*	.05*	–	.05

(20)

0	.05*	–	–
.05*	–	.05*	–
.05*	–	.05*	–
–	–	–	.05*

Out of these 20 feasible samples 15 are distinct.

The variance of $\bar{\bar{y}}$, where $\bar{\bar{y}}$ is $\hat{Y}/M$, the estimated total of y in the frame, will be weighted S^2_{RC} and the weights will depend on the manner in which the feasible samples are drawn. Since the manner of selecting the feasible samples depends upon the sampler, no mathematical expression for Var ($\bar{\bar{y}}$) can be given at this stage. However, along the lines of Jessen (1975), a nearly unbiased estimate of variance of $\bar{\bar{y}}$ can be obtained by using the expression

$$\hat{Var}(\bar{\bar{y}}) = [1 - n\sum_{i=1}^{L}\sum_{j=1}^{L}(M_{ij}/M)^2](\tilde{s}^2_{RC}/n)$$

where $\tilde{s}^2_{RC}$ is the pooled estimate for the interaction S^2_{RC} and can be obtained along the lies of Section 2 for each of the feasible samples, after re-arrangement of the feasible samples in the form of m latins along the diagonal and using the expression similar to that of $\hat{S}^2_{RC}$. To illustrate, the first and the second feasible samples can be considered as randomized forms of the following latin arrangements.

(1)

	1	4	2	3
2	A	–	–	–
3	B	A	–	–
1	–	–	A	B
4	–	–	B	–

(2)

	2	3	1	4
1	A	B	–	–
4	B	–	–	–
2	–	–	A	–
3	–	–	B	A

For the two-way frame discussed above, the estimates for the variance of $\bar{\bar{y}}$ for t^{th} ($t = 1, ..., 20$) feasible sample, denoted by $[\hat{Var}(\bar{\bar{y}})]_t$ are computed as

Sample No. (t)	$[\hat{Var}(\bar{\bar{y}})]_t$	Sample No. (t)	$[\hat{Var}(\bar{\bar{y}})]_t$
1	0.0066120	11	0.0169422
2	0.0066120	12	0.0069051
3	0.0066120	13	0.0019048
4	0.0066120	14	0.0761944
5	0.0077842	15	0.0011172
6	0.0000000	16	0.0104217
7	0.0297634	17	0.0206054
8	0.0297634	18	0.0048537
9	0.0728426	19	0.0140849
10	0.0206054	20	0.0002472

The probability of selection (P_t) for each of these 20 feasible samples is 0.05. Therefore

$$E[\hat{Var}(\bar{\bar{y}})] = \sum_{t=1}^{20} [\hat{Var}(\bar{\bar{y}})]_t . P_t = 0.0170236$$

A similar approach may also be taken for the case of R×C rectangular population when $n \neq r.t.$

It may be mentioned here that the solution to the above two-way problem for n = 8, based on adaptive method, is reported by Jessen (1978). There seems to be an error in his solution for the second feasible sample where $n.A_i$ is 0.4 and not 0.2, as taken by him. If we take $n.A_i$ as 0.4 and take the same designated cells as in Jessen, the method fails to generate the third feasible sample, with the condition that exactly two cells are to be taken from each row and column.

REFERENCES

Bryant, E.C., Hartley, H.O. and Jessen, R.J. (1960). Design and estimation in two-way stratification. *J. Amer. Statist. Assoc.*, **55**, 105-124.

Causey, B.D., Cox, L.H. and Ernst, L.R. (1985). Applications of transportation theory to statistical problems. *J. Amer. Statist. Assoc.*, **80**, 903-909.

Cox, Lawrence H. (1987). A constructive procedure for unbiased controlled rounding. *J. Amer. Statist. Assoc.*, **82**, 520-524.

Ernst, Lawrence R. (1981). A constructive solution for two-dimensional controlled selection problems. *Proceedings of the Survey Research Methods Section, Amer. Statist. Assoc., Washington, D.C.*, 61-64.

Goodman, R. and Kish, L. (1950). Controlled selection - A technique in probability sampling. *J. Amer. Statist. Assoc.*, **45**, 350-372.

Hess, I. and Srikanthan, K.S. (1966). Some aspects of the probability sampling technique of controlled selection. *Health Serv. Res.*, **Summer 1966**, 8-52.

Jessen, R.J. (1970). Probability sampling with marginal constraints. *J. Amer. Statist. Assoc.*, **65**, 776-796.

Jessen, R.J. (1973). Some properties of probability lattice sampling. *J. Amer. Statist. Assoc.*, **68**, 26-28.

Jessen, R.J. (1975). Square and cubic lattice sampling. *Biometrics*, **31**, 449-471.

Jessen, R.J. (1978). *Statistical Survey Techniques*. John Wiley & Sons, New York.