数理解析研究所講究録
396 巻 1980 年 260-288

260

MULTIDIMENSIONAL SYSTEMATIC SAMPLING

Tateaki Sasaki

The Institute of Physical and Chemical Research
Wako-shi, Saitama 351, Japan

ABSTRACT

A desirable sampling should be such that each element in the population is treated evenly and elements sampled are distributed uniformly in the population. Furthermore, a useful sampling should be such that we can apply various accuracy increasing tricks easily. This paper proposes a multidimensional sampling method which possesses these properties. The method is based on an analogue of counting numbers of d figures and an elementary property of prime numbers, and it samples elements systematically. The method is formulated so as to accommodate with sampling with unequal probability. Various merits of the method are pointed out and numerical investigations are given.

KEY WORDS AND PHRASES: multidimensional sampling, sampling theory, accuracy increasing technique, cluster sampling, systematic sampling, sampling with unequal probability, evenness of sampling, uniformity of the distribution, Monte Carlo method.
CR CATEGORIES: 4.49, 4.9, 5.12, 5.16, 5.5.

§1. Introduction

Many important sampling methods were developed before 1950, many of which can be found in reference [1] or [2]. Such methods were designed mostly from the need for sample surveys of events in our world, such as opinions of the people to the national topics, health status of the people, the number of fishes in a lake, and so on. On the other hand, computer sampling from populations in the mathematical world becomes more and more important today. This paper proposes a multidimensional sampling method which is suited for executing by a computer and for using together with various accuracy increasing tricks. Since multidimensional mathematical populations often have large variances, the use of some accuracy increasing trick is inevitably necessary in actual applications.

It is well recognized that a key for increasing accuracy of the estimation of a mean $\bar{X}$ is to utilize informations on the variate X in the population [3], [4], [5], [6]. Using such an information, we can perform effective sampling with unequal probability. Note that another sampling is necessary for obtaining such informations. Actually, sampling is performed in several phases, and information obtained in earlier phases are utilized to improve the sampling probability function for later phases. A desirable sampling method is, therefore, such that the necessary informations are obtained most.

In the case of sampling with equal probability, the desirable method is such that any element in the population has an equal probability of being sampled and the sampled elements are distributed in the population as uniformly as possible. The simple random sampling satisfies the first condition, but the distribution of sampled elements is not so uniform. The distribution is much more uniform in stratified sampling than in random sampling. The applicability of stratified sampling is, however, restricted severely by the constraint that the size of a sample is not less than two times of the number of strata. The method proposed in this paper is a systematic sampling which is applicable even in high dimensions. It is important to note that various accuracy increasing tricks are conveniently and effectively used in our method because of its systematicness.

Several authors investigated multidimensional systematic sampling

about thirty years ago (see, reference [1], §8.13). It is interesting to comment that Homeyer and Black's and Yates and Patterson's methods [7], [8], [9], based on the latin square have a similarity to ours, although their design principles are different from ours and their methods are less elegant than ours.

In §2, we briefly survey three sampling methods, simple random sampling, cluster sampling, and systematic sampling, for finite populations in one dimension. The uniformity of distribution of the sampled elements is defined in §3 so as to accommodate with sampling with unequal probability, and a multidimensional systematic sampling is proposed. Detailed anlysis of our method is presented in §4. Applicability of various accuracy increasing tricks in our method is explained in §5. Various merits of the method are also pointed out. Results of numerical investigations of the method are reported, too.

§2. Survey of sampling methods

We present in this section some elementary results of sampling theories for finite populations in one dimension. For details, see reference [1] or [2].

2.1. Simple random sampling

Let $\mathfrak{Y}$ denote a population of N elements $y_1, y_2, \cdots, y_N$:

$$\mathfrak{Y}: \{y_1, y_2, \cdots, y_N\}.$$

The population mean is denoted by $\bar{Y}$ and defined as

$$\bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i. \tag{1}$$

Let a sample of $\mathfrak{Y}$ be $\mathfrak{S}$, which is composed of n elements $y_1, y_2, \cdots, y_n$ sampled randomly from $\mathfrak{Y}$:

$$\mathfrak{S}: \{y_1, y_2, \cdots, y_n\}.$$

The sample mean is denoted by $\bar{y}$ and defined as

$$(2) \qquad \bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i .$$

The $\bar{y}$ is an unbiased estimator of $\bar{Y}$, that is

$$E(\bar{y}) = \bar{Y},$$

where E denotes averaging over all possible samples of size n taken from the population $\mathfrak{P}$.

The sample mean $\bar{y}$ is deviated from the population mean $\bar{Y}$ by some amount. A useful criterion to represent this deviation is the variance, or the mean square deviation, of $\bar{y}$:

$$(3) \qquad V(\bar{y}) \equiv E((\bar{y}-\bar{Y})^2) = \frac{1}{n} \frac{N-n}{N-1} \sigma^2,$$

where σ^2 is the population variance defined as

$$(4) \qquad \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i-\bar{Y})^2 .$$

A slightly different definition of the variance is

$$(5) \qquad S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i-\bar{Y})^2, \quad \text{or} \quad S^2 = \frac{N}{N-1} \sigma^2 .$$

In terms of S^2, the variance of $\bar{y}$ is given as

$$(6) \qquad V(\bar{y}) = \frac{S^2}{n} (1-\frac{n}{N}) \equiv \frac{S^2}{n} (1-f),$$

where f is called the finite population correction. The quantity

$$(7) \qquad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i-\bar{y})^2$$

is an unbiased estimator of S^2. Therefore, the sample $\mathfrak{S}$ allows us to calculate not only an approximate value of $\bar{Y}$ but also an estimate of its error, $(s/\sqrt{n})\sqrt{1-f}$.

2.2. Cluster sampling

Suppose all the elements of $\mathcal{Y}$ are divided into M subpopulations of sizes $N_1, N_2, \cdots, N_M$, respectively:

$$\mathcal{Y}: \{\{y_{11}, y_{12}, \cdots, y_{1N_1}\}, \cdots, \{y_{M1}, y_{M2}, \cdots, y_{MN_M}\}\},$$

$$N = N_1 + N_2 + \cdots + N_M.$$

These subpopulations are called *clusters*. In cluster sampling, a sample $\mathcal{S}$ of $\mathcal{Y}$ is constructed by first sampling m clusters randomly and then sampling elements randomly from each sampled clusters. Suppose the first m clusters are sampled and n_i elements are sampled from the i-th cluster:

$$n = n_1 + n_2 + \cdots + n_m.$$

An unbiased estimator of $\bar{Y}$ is given by

$$(8) \qquad \bar{y}_{cl} \equiv \frac{1}{m}\sum_{i=1}^{m}\left(\frac{MN_i}{N}\right)\bar{y}_i, \quad \bar{y}_i \equiv \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}.$$

Here, $\bar{y}_i$ is the sample mean for the i-th cluster. Note that N/M is the average size of the clusters.

The variance of $\bar{y}_{cl}$ can be calculated by noting that the sampling is performed in two stages, the first is the cluster selection and the second is the element selection in each selected cluster. According to the elementary theory of probability, the variance of $\bar{y}$ in two-stage sampling is given by

$$V(\bar{y}) = V(E(\bar{y}|\nu_1,\cdots,\nu_m)) + E(V(\bar{y}|\nu_1,\cdots,\nu_m)).$$

Here, $E(\bar{y}|\nu_1,\cdots,\nu_m)$ is the conditional mean of $\bar{y}$, or the expectation of $\bar{y}$ over clusters $\nu_1, \nu_2, \cdots, \nu_m$, and $V(\bar{y}|\nu_1,\cdots,\nu_m)$ is the conditional variance of $\bar{y}$. The first term of the above equation gives

$$V\left(\frac{1}{m}\sum_{i=1}^{m}\left(\frac{MN_{\nu_i}}{N}\right)\bar{y}_{\nu_i}\right) = \frac{S_b^2}{m}\left(1-\frac{m}{M}\right), \quad S_b^2 = \frac{1}{M-1}\sum_{i=1}^{M}\left(\frac{MN_i}{N}\bar{y}_i-\bar{Y}\right)^2.$$

The second term gives

$$E\left(\frac{1}{m^2}\sum_{i=1}^{m}\left(\frac{MN_{\nu_i}}{N}\right)^2\frac{S_{\nu_i}^2}{n_{\nu_i}}\left(1-\frac{n_{\nu_i}}{N_{\nu_i}}\right)\right)=\frac{1}{mM}\sum_{i=1}^{M}\left(\frac{MN_i}{N}\right)^2\frac{S_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right),$$

where

$$S_i^2=\frac{1}{N_i-1}\sum_{j=1}^{N_i}(y_{ij}-\bar{Y}_i)^2 \quad \text{with} \quad \bar{Y}_i\equiv\frac{1}{N_i}\sum_{j=1}^{N_i}y_{ij}.$$

Therefore,

$$(9)\qquad V(\bar{y}_{cl})=\frac{1}{m}\left\{S_b^2\left(1-\frac{m}{M}\right)+\frac{1}{M}\sum_{i=1}^{M}\left(\frac{MN_i}{N}\right)^2\frac{S_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)\right\}.$$

The S_b^2 is the variance of the cluster mean and S_i^2 is the variance of the element in the i-th cluster. An unbiased estimator of $V(\bar{y}_{cl})$ is

$$(10)\qquad v(\bar{y}_{cl})=\frac{1}{m}\left\{s_b^2\left(1-\frac{m}{M}\right)+\frac{1}{m}\sum_{i=1}^{m}\left(\frac{MN_i}{N}\right)^2\frac{s_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)\right\},$$

where

$$(11)\qquad s_b^2=\frac{1}{m-1}\sum_{i=1}^{m}\left(\frac{MN_i}{N}\bar{y}_i-\bar{y}_{cl}\right)^2 \quad \text{and} \quad s_i^2=\frac{1}{n_i-1}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_i)^2.$$

2.3. Systematic sampling

Suppose the size of the population is a multiple of the size of the sample, $N = Mn$, and all the elements of $\mathcal{Y}$ are systematically divided into M clusters in the following way:

$$\mathcal{Y}: \{\{y_1, y_{M+1}, \cdots, y_{(n-1)M+1}\}, \{y_2, y_{M+2}, \cdots, y_{(n-1)M+2}\},$$
$$\cdots\cdots\cdots\cdots, \{y_M, y_{M+M}, \cdots, y_{(n-1)M+M}\}\}$$
$$\equiv \{\{y_{11}, y_{12}, \cdots, y_{1n}\}, \cdots, \{y_{M1}, y_{M2}, \cdots, y_{Mn}\}\}.$$

In systematic sampling, a randomly chosen cluster is used as a sample of $\mathcal{Y}$. Suppose the i-th cluster is chosen as a sample, then the sample mean $\bar{y}_i$ is an unbiased estimator of $\bar{Y}$:

$$(12) \qquad \bar{y}_{sy} \equiv \bar{y}_i = \frac{1}{n}\sum_{j=1}^{n} y_{ij}.$$

The variance of $\bar{y}_{sy}$ is given by

$$(13) \qquad V(\bar{y}_{sy}) = \frac{1}{M}\sum_{i=1}^{M}(\bar{y}_i-\bar{Y})^2 = \frac{\sigma^2}{n}\{1 + (n-1)\rho_w\},$$

$$\rho_w = \frac{1}{\sigma^2}\frac{1}{N(n-1)}\sum_{i=1}^{M}\sum_{j\neq k}^{n}(y_{ij}-\bar{Y})(y_{ik}-\bar{Y}).$$

The ρ_w is called the intracluster correlation coefficient, and it satisfies the relation $-1/(n-1) \leqq \rho_w \leqq 1$.

Formula (13) shows that the accuracy of the estimator $\bar{y}_{sy}$ depends strongly on the correlation coefficient ρ_w, i.e., on how the elements of $\mathfrak{Y}$ are labeled. In some cases, where $\rho_w < 0$, systematic sampling will give more accurate answers than simple random sampling. In many cases, where elements in populations may be considered to be labeled randomly, systematic sampling will give almost the same results as random sampling. If, however, the elements in the population have a periodic trend with a period of a multiple or a factor of M, systematic sampling will give less accurate answer than random sampling.

It is important to note that, if we choose only one systematic cluster as a sample, we have no unbiased estimator of $V(\bar{y}_{sy})$. If we need an estimation of the error of $\bar{y}_{sy}$, we may choose m clusters as a sample and estimate $\bar{Y}$ by the estimator

$$(14) \qquad \bar{y}_{sy} = \frac{1}{m}\sum_{i=1}^{m}\bar{y}_i.$$

Then, an unbiased estimator of $V(\bar{y}_{sy})$ is given by

$$(15) \qquad v(\bar{y}_{sy}) = \frac{s_b^2}{m}(1-\frac{m}{M}), \quad s_b^2 = \frac{1}{m-1}\sum_{i=1}^{m}(\bar{y}_i-\bar{y}_{sy})^2.$$

This formula is equivalent to (10) with $n_i = N_i = n/m$, $i=1,2,\cdots,m$.

§3. Proposal of a multidimensional systematic sampling

3.1. Multidimensional population

Let $\mathcal{Y}$ denote a population in which each element y is labeled by parameters x_1, x_2, $\cdots$, and x_d:

$$\mathcal{Y}: \{y(x_1,x_2,\cdots,x_d)\}.$$

Let us assume that the parameter range is normalized as

$$0 \leqq x_i \leqq 1, \quad i=1,2,\cdots,d.$$

If all the parameters are continuous the population mean $\bar{Y}$ and the population variance V(y) are

$$\bar{Y} = \int_0^1 \cdots \int_0^1 y(x_1,\cdots,x_d)\, dx_1\cdots dx_d,$$

$$V(y) = \int_0^1 \cdots \int_0^1 y^2(x_1,\cdots,x_d)\, dx_1\cdots dx_d - \bar{Y}^2,$$

and our problem is nothing but the multidimensional integration and the estimation of its error. The population $\mathcal{Y}$ may be of finite-sized, but we assume the size is so large that we may regard the parameters as almost continuous.

We divide the range [0,1] for the parameter x_i into p_i subranges:

$$\{[x_i^{(0)},x_i^{(1)}], (x_i^{(1)},x_i^{(2)}], \cdots, (x_i^{(p_i-1)},x_i^{(p_i)}]\},$$

where

$$0 = x_i^{(0)} < x_i^{(1)} < x_i^{(2)} < \cdots < x_i^{(p_i-1)} < x_i^{(p_i)} = 1.$$

The population $\mathcal{Y}$ is divided into p_i subpopulations by this division. We call such a subpopulation *a stratum w.r.t.* x_i. We perform the above division for each parameter. Then, the parameter space which is the d-dimensional unit hyper-cube is divided into small nonoverapping hyper-rectangles the number of which is $p_1p_2\cdots p_d$. We call such a hyper-rectangle *a cell*. We assume the number of cell is much larger than the size of a sample.

3.2. Definitions of uniformity and evenness

In considering the error in a quasi-Monte Carlo method, Koksma [10] introduced the concept of discrepancy of a set of points in the one-dimensional interval [0, 1], which gives a measure of how uniformly the points are distributed. The concept was generalized to the d-dimensions by Hlawka [11]. The discrepancy is defined in the following way. Let R be a hyper-rectangle consisting of all points $(x_1,x_2,\cdots,x_d)$ satisfying

$$a_i \leqq x_i \leqq b_i, \quad i=1,2,\cdots,d,$$

here the numbers a_i and b_i satisfy the conditions

$$0 \leqq a_i < b_i \leqq 1, \quad i=1,2,\cdots,d.$$

Let $|R|$ denote the volume of R:

$$|R| \equiv (b_1-a_1)(b_2-a_2)\cdots(b_d-a_d).$$

Let n points be sampled from the d-dimensional unit hyper-cube and let $Z_n(R)$ be the number of points sampled from R. The discrepancy of these n points is the least upper bound of the quantity

$$(16) \qquad |Z_n(R)/n - |R||$$

over all the subsets R contained in the unit hyper-cube.

According to this definition, the discrepancy of points sampled by the square-grid sampling method (cf. Fig.1-a) is quite large: suppose we sample all mesh points of $\{x_i=(j-0.5)/m \mid j=1,2,\cdots,m;\ i=1,2,\cdots,d\}$ then the number of points sampled is m^d while their discrepancy is as large as $1/m$, since no point is sampled from domains $\{(x_1,x_2,\cdots,x_d) \mid (j-0.5)/m<x_1<(j+0.5)/m;\ 0\leqq x_i\leqq 1,\ i=2,3,\cdots,d\}$, $j=1,2,\cdots,m-1$, and these domains are of volume $1/m$. This large discrepancy is contradictory to our common sense: our common sense is that square-grid sampling gives points distributed uniformly. The contradiction arises from that R is a hyper-rectangle and not a hyper-sphere. We, therefore, introduce another definition of uniformity. The definition is stated so as to

be applied to sampling with unequal probability.

We denote the cell surrounded by 2d hyper-planes $x_i = x_i^{(j_i-1)}$ and $x_i = x_i^{(j_i)}$, i=1,2,···,d, by

$$C[j_1,j_2,\cdots,j_d].$$

Definition 1. *The distance $\rho(c,c')$ between cells $c = C[j_1,j_2,\cdots,j_d]$ and $c' = C[j'_1,j'_2,\cdots,j'_d]$ is*

$$\rho(c,c') = \{(\frac{j_1-j'_1}{p_1})^2 + (\frac{j_2-j'_2}{p_2})^2 + \cdots + (\frac{j_d-j'_d}{p_d})^2\}^{1/2}. \tag{17}$$

We call a set of all cells that are located within a distance ρ from a given cell a sphere of radius ρ.

Definition 2. *Let n cells be sampled, and extend the sampled cells periodically: if $C[j_1,j_2,\cdots,j_d]$ is sampled, then regard all the cells $C[j_1+i_1p_1,j_2+i_2p_2,\cdots,j_d+i_dp_d]$ with arbitrary integers i_1, i_2, ···, i_d as being sampled. Consider a sphere R and let the volume of R be $|R|$. Let $z_n(R)$ denote the number of cells sampled from R. The discrepancy of these n cells is the least upper bound of the quantity*

$$|z_n(R)/n - |R||. \tag{18}$$

Definition 3. *The distribution of a set of cells is uniform if the discrepancy of them is of the same order as the discrepancy of almost the same number of cells sampled by square-grid sampling.*

Similarly to the definition of uniformity, we define the evenness so as to accommodate with sampling with unequal probability.

Definition 4. *If all the cells have an equal probability of being sampled, then the sampling is even.*

3.3. Requirements on sampling

We require our sampling method to be even and to give a uniform distribution in a stronger sense than that given in definition 3.

i) Sampling must be even.

ii) Sampling must give a highly uniform distribution such that the sampled elements are highly evenly distributed over all strata w.r.t. each x_i.

iii) Sampling must be performed fast and easily.

Let us explain the second requirement by considering a simple example in two dimension: let p_1 and p_2 be 13 so that the parameter space is divided into 169 cells, and let us sample about 25 elements. The elements obtained by square-grid sampling are distributed uniformly, as is illustrated by Fig.1-a. We see, however, that the sampled elements are quite unevenly distributed over the strata w.r.t. x_1 and x_2. Let us next consider the sample illustrated by Fig.1-2, where the elements located at the center of cells numbered are sampled in this order. We can see that the sampled elements are highly uniformly distributed globally as well as over the strata w.r.t. x_1 and x_2. Note that the sample in Fig.1-b is systematic: the sampling was performed parallelly with the x_2 axis by skipping six cells in the x_2 direction and two strata in the x_1 direction. If we skip no stratum in the x_1 direction, we have the sample illustrated by Fig.1-c, the cells of which are not uniformly distributed.

| Fig.1-a | | Fig.1-b | | Fig.1-c |

3.4. Algorithm

The sampling method to be proposed in this paper is a straightforward generalization of the method illustrated by Fig.1-b. The essential factor that makes our sampling method possible is that all p_i, $i=1,2,\cdots,d$, are chosen to be prime numbers. Since we assume the number of cells is sufficiently large, our problem is essentially how to sample cells, and how to sample an element in each sampled cell is a subsidiary problem. We, therefore, describe only the method for sampling cells.

Let $p_1, p_2, \cdots, p_d$ be prime numbers, which are not necessarily different from each other. Suppose we want to sample about n_c cells, where

$$\max[p_1, p_2, \cdots, p_d] \ll n_c \ll p_1 p_2 \cdots p_d.$$

Let $k_1, k_2, \cdots, k_d$ denote positive integers such that

$$0 < k_1 < p_1 \quad \text{(detailed specification is given in §4.2)},$$

$$p_2/k_2 \simeq p_3/k_3 \simeq \cdots \simeq p_d/k_d, \tag{19}$$

$$p_1 p_2 \cdots p_d/(k_2 k_3 \cdots k_d) \simeq n_c. \tag{20}$$

We give the term *leading cells* to the following set of cells

$$\{C[1,j_2,\cdots,j_d] \mid 1 \leqq j_2 \leqq k_2, \cdots, 1 \leqq j_d \leqq k_d\}.$$

The shaded area in Fig.1-b manifests the leading cells. It is easy to determine $k_2, k_3, \cdots, k_d$ satisfying (19) and (20). Our sampling is performed by the following algorithm:

Algorithm SYSCLUST

Being given a set of prime numbers $(p_1,p_2,\cdots,p_d)$ and a set of increments $(k_1,k_2,\cdots,k_d)$ satisfying (19) and (20), this algorithm samples about n_c cells systematically.

Step 1. [Initial cell.] Select an initial cell randomly from the leading cells.

Step 2. [Next cell.] Sample the next cell according to the following procedure:

```
PROCEDURE NEXTCELL([j1,j2,···,jd])
%comment: this procedure determines a cell to be
          sampled next to the cell C[j1,j2,···,jd];
          i ← d;
INCREASE: j ← ji+ki;
          if j > pi then go to UPPLACE;
          replace ji in [j1,j2,···,jd] by j;
          return [j1,j2,···,jd];
 UPPLACE: replace ji in [j1,j2,···,jd] by j-pi;
          if i = 1 then return [j1,j2,···,jd];
          i ← i-1;
          go to INCREASE;
          end;
```

Step 3. [Terminate.] If all strata w.r.t. x_1 are scanned then terminate sampling else go to Step 2.

That the above algorithm terminates is easily proved by the following well-known fact in elementary number theory:

$$\{a, a+b, a+2b, \cdots, a+(p-1)b\} \equiv \{0, 1, 2, \cdots, p-1\} \pmod{p},$$

which is valid for any integers a, $b \neq 0$, and prime number p. Similarly, it can be easily proved that the cells sampled are highly evenly distributed over all the strata w.r.t. each x_i. Furthermore, the algorithm is quite fast and easily programmed. Detailed analysis of the algorithm is given in the next section.

Let us show the performance of the above algorithm by a simple example in three dimension, where $p_1=p_2=p_3=7$, $k_1=k_2=k_3=2$, and the initial cell C[1,1,2]. Then, the sampling will be done as follows:

$$\begin{aligned} &C[1,1,2] \to C[1,1,4] \to C[1,1,6] \to C[1,3,1] \to C[1,3,3] \\ &\to C[1,3,5] \to C[1,3,7] \to C[1,5,2] \to C[1,5,4] \to C[1,5,6] \\ &\to C[1,7,1] \to C[1,7,3] \to C[1,7,5] \to C[1,7,7] \to C[3,2,2] \\ &\to \cdots. \end{aligned}$$

The sampling will terminate after scanning the first, third, fifth, seventh, second, fourth, and sixth strata w.r.t. x_1 in this order. We can see the sampling is similar to counting numbers modulo 7^3, where the numbers are of three figures with radix 7 and the increment in each place is 2.

§4. Analysis of the sampling method proposed

Let N and M denote the number of total cells and the number of leading cells, respectively:

$$N = p_1 p_2 \cdots p_d, \tag{21}$$

$$M = k_2 k_3 \cdots k_d. \tag{22}$$

Let us call a set of cells sampled by algorithm SYSCLUST *a cluster*.

4.1. Structure of a cluster

It is clear that algorithm SYSCLUST samples only one leading cell when it is scanning the first stratum w.r.t. x_1, and no leading cell is contained in other strata w.r.t. x_1. The algorithm will scan p_1 strata w.r.t. x_1 successively, and the scanning terminates before the first stratum is scanned again. This proves the following lemma:

Lemma 1. *Algorithm SYSCLUST samples only one leading cell, that is the cell sampled initially.*

Let us next consider what happens when we remove the termination condition in algorithm SYSCLUST. Then, the sampling will continue forever, while the number of cells is finite. Hence, the same cell will be sampled periodically. Suppose a cell is sampled twice just after sampling ℓ_d cells. This means that the sampling is done by transferring from a stratum w.r.t. x_d to another ℓ_d times, $\ell_{d-1} = [\ell_d k_d / p_d]$ times for the strata w.r.t. x_{d-1}, $\ell_{d-2} = [\ell_{d-1} k_{d-1} / p_{d-1}]$ times for the strata w.r.t. x_{d-2}, and so on, where $[\alpha]$ is the Gauss' notation denoting the largest integer not exceeding α. That is

$$\ell_i k_i = \ell_{i-1} p_i + r_i, \quad 0 \leqq r_i < p_i, \quad i=2,3,\cdots,d,$$

$$\ell_1 k_1 = \ell_0 p_1 + r_1, \quad 0 \leqq r_1 < p_1.$$

On the other hand, since the same cell is sampled again, $\ell_d k_d$, $\ell_{d-1} k_{d-1}$, $\cdots$, $\ell_1 k_1$ must be multiples of p_d, p_{d-1}, $\cdots$, p_1, respectively:

$$\ell_i k_i = q_i p_i, \quad q_i \text{ is an integer}, \quad i=1,2,\cdots,d.$$

These relations on ℓ_i are consistent only when $r_i = 0$ and $q_i = \ell_{i-1}$:

$$\ell_1 k_1 = q_1 p_1,$$
$$\ell_2 k_2 = \ell_1 p_2,$$
$$\cdots\cdots$$
$$\ell_d k_d = \ell_{d-1} p_d.$$

The first two of these equalities give $\ell_2 k_1 k_2 = q_1 p_1 p_2$. Since

$GCD(k_1k_2, p_1p_2) = 1$, ℓ_2 must be a multiple of p_1p_2. Similarly,

$$\ell_i k_1 k_2 \cdots k_i = q_1 p_1 p_2 \cdots p_i, \quad i=2,3,\cdots,d,$$

and ℓ_i must be a multiple of $p_1p_2\cdots p_i$, $i=1,2,\cdots,d$. Since the number of total cells is $p_1p_2\cdots p_d$, the same cell is sampled if $\ell_d = p_1p_2\cdots p_d$. This gives $q_1 = k_1k_2\cdots k_d$ and $\ell_i = p_1p_2\cdots p_i k_{i+1}k_{i+2}\cdots k_d$, $i=1,2,\cdots,d$. Thus, we have proved the following lemma:

Lemma 2. *The procedure NEXTCELL in algorithm SYSCLUST does not sample the same cell until it samples all the N cells.*

If $C[j_1,j_2,\cdots,j_d]$ is the cell sampled last by algorithm SYSCLUST, then it satisfies the conditions

$$j_1 \equiv 1 + (p_1-1)k_1 \pmod{p_1},$$
$$p_i - k_i < j_i \leqq p_i, \quad i=2,3,\cdots,d.$$

Hence, the cell to be sampled next to $C[j_1,j_2,\cdots,j_d]$ by procedure NEXTCELL in algorithm SYSCLUST must be one of the leading cells. This consideration leads us to the following corollaries:

Corollary 1. *Any two sets of cells sampled by algorithm SYSCLUST with two different initial cells have no common cell.*

Corollary 2. *The total N cells are completely divided into M non-overapping clusters by algorithm SYSCLUST.*

Since a cluster is selected randomly in algoritm SYSCLUST, each cell has an equal probability of being sampled due to corollary 2. This proves the evenness of our sampling. Since the number of clusters is equal to the number of leading cells, the average number of cells contained in one cluster is

$$N/M = p_1p_2\cdots p_d/(k_2k_3\cdots k_d) \simeq n_c.$$

This proves that algorithm SYSCLUST samples about n_c cells.

4.2. Uniformity of the distribution

Let us first consider the minimum distance between two sampled cells c and c' which are contained in the same stratum w.r.t. x_1. Suppose c and c' are not contained in the same stratum w.r.t. x_2. Then, the distance is not less than k_2/p_2 because these cells are sampled with the increment k_2 in the x_2 direction. Next, suppose the cells are contained in the same stratum w.r.t. x_2 but not in the same stratum w.r.t. x_3. Then, the distance is not less than k_3/p_3 by the similar reason to the above. Continuing this reasoning, we find the following lemma:

Lemma 3. *Let a cell* $c = C[j_1, j_2, \cdots, j_d]$ *be sampled by algorithm SYSCLUST. Then, no cell is sampled from the set of cells* $\{C[j_1, j_2', \cdots, j_d'] \mid j_i \leq j_i' < j_i + k_i,\ i=2,3,\cdots,d\}$ *except for c.*

Corollary 3. *In each stratum w.r.t.* x_1, *the distance between two cells sampled by algorithm SYSCLUST is not less than*

$$\rho_1 = \min[k_2/p_2,\ k_3/p_3,\ \cdots,\ k_d/p_d].$$

If we consider the distance between cells contained in different strata w.r.t. x_1, the minimum distance may be less than ρ_1.

Let us next consider the maximum distance between two neighboring cells sampled. Let us consider, for a given cell $C[j_1, j_2, \cdots, j_d]$, the set of cells $\{C[j_1, j_2', j_3', \cdots, j_d'] \mid j_2 \leq j_2' < j_2 + k_2;\ 1 \leq j_i' \leq p_i,\ i=3,4,\cdots,d\}$. Since the increment in the x_2 direction is k_2, algorithm SYSCLUST samples at least one cell (for some value of j_2' from this set), let the value be $\tilde{j}_2$, and does not sample cell for other value of j_2'. Next, consider the set of cells $\{C[j_1, \tilde{j}_2, j_3', \cdots, j_d'] \mid j_3 \leq j_3' < j_3 + k_3;\ 1 \leq j_i' \leq p_i,\ i=4,5,\cdots,d\}$, then we see algorithm SYSCLUST samples cells from this set for only one value of j_3' by the similar reason to the above. Continuing this reasoning, we finally see that algorithm SYSCLUST samples at least one cell from the set $\{C[j_1, \tilde{j}_2, \cdots, \tilde{j}_{d-1}, j_d'] \mid j_d \leq j_d' < j_d + k_d\}$. Since the increment in the x_d direction is k_d, it is impossible to sample more than one cell from this set. This leads us to the following lemma:

Lemma 4. *For any cell* $C[j_1,j_2,\cdots,j_d]$, *algorithm SYSCLUST samples only one cell from the set of sells* $\{C[j_1,j_2',\cdots,j_d'] \mid j_i \leq j_i' < j_i+k_i, i=2,3,\cdots,d\}$.

Corollary 4. *The maximum of the distance between two neighboring cells sampled by algorithm SYSCLUST is not greater than*

$$\rho_2 = \{(k_2/p_2)^2 + (k_3/p_3)^2 + \cdots + (k_d/p_d)^2\}^{1/2}.$$

Lemmas 3 and 4 show that, in each stratum w.r.t. x_1, the distribution of the cells sampled by our method is quite similar to that in square-grid sampling. The corollary 4 tells us that the discrepancy of the cells sampled by algorithm SYSCLUST is of order

$$D_1 \equiv \rho_2^d \simeq d^{d/2}\langle k/p\rangle^d, \tag{23}$$

where $\langle k/p\rangle$ is the average of k_i/p_i, i=2,3,···,d:

$$\langle k/p\rangle = (k_2/p_2 + k_3/p_3 + \cdots + k_d/p_d)/(d-1).$$

Using the relation (20), we have $1/n_c \simeq \langle k/p\rangle^{d-1}/p_1$ and

$$D_1 \simeq d^{d/2}p_1\langle k/p\rangle/n_c. \tag{23'}$$

On the other hand, if about n_c cells are sampled by the square-grid sampling methd, the minimum distance between two cells sampled is about

$$\rho_0 = (1/n_c)^{1/d},$$

and the maximum of the distance between two neighboring cells sampled is about

$$(\rho_0^2 + \rho_0^2 + \cdots + \rho_0^2)^{1/2} = d^{1/2}\rho_0.$$

Hence, the discrepancy of the cells sampled by the square-grid sampling method is of order

$$D_0 \equiv (d^{1/2}\rho_0)^d = d^{d/2}/n_c. \tag{24}$$

The quantity D_1 is not so much different from D_0, and algorithm SYSCLUST gives a highly uniform distribution. Note that, if we sample cells randomly, the discrepancy of the cells sampled is $O(1/\sqrt{n_c})$ and it is much greater than D_0 and D_1.

So far, we have not mentioned on how to determine the value of k_1. (The above discussions are valid for any value of k_1.) Comparison of Figs.1-b and 1-c shows that some value of k_1 will give a highly uniform distribution in the x_1-x_2 plane while another value of k_1 will not. We want to determine the value of k_1 so that the discrepancy of the cells sampled may be minimized. The determination of the best value of k_1 is, however, quite difficult. We, therefore, calculate only a good value of k_1, instead of the best value, in the following way. We project the d-dimensional distribution on the x_1-x_2 plane and consider only the two-dimensional distribution, as is shown by Fig.1. Note that the projected distribution is the same as that obtained by applying algorithm SYSCLUST in two dimension. Then, for a given k_1, we can easily compute the minimum of the two-dimensional distance between neighboring cells sampled. By changing k_1 from 1 to $[p_1/2]$, we determine the value of k_1 so that the minimum is maximized. This process can be performed easily by a computer. The distribution shown in Fig.1-b was obtained in this way.

4.3. The mean and the error

We have seen that the total N cells are divided into M nonoverlapping clusters. In actual sampling, we may apply algorithm SYSCLUST several times and sample several clusters. Suppose we sample m different clusters randomly. Let n_i denote the number of cells contained in the i-th cluster, W_i the size of the i-th cluster, and w_{ij} the size of the j-th cell in the i-th cluster:

$$N = \sum_{i=1}^{M} n_i,$$

$$W_i = \sum_{j=1}^{n_i} w_{ij}.$$

Let $\bar{y}_i$ be an unbiased estimator of the i-th cluster mean $\bar{Y}_i$. The expression of $\bar{y}_i$ is the same as that for stratified sampling, since in both our and stratified sampling methods, the population is divided into subpopulations (in our case, a cluster is divided into cells) and sampling is done in each subpopulation. That is,

$$(25) \qquad \bar{y}_i = \sum_{j=1}^{n_i} (w_{ij}/W_i)\, \bar{y}_{ij},$$

where $\bar{y}_{ij}$ is a sample mean in the j-th cell of the i-th cluster. According to section 2.2, an unbiased estimator of the total population mean $\bar{Y}$ is given by

$$(26) \qquad \bar{y} = \frac{1}{m} \sum_{i=1}^{m} (MW_i)\, \bar{y}_i.$$

Next, let us consider the variance of $\bar{y}$, the square root of which gives an estimate of the error of $\bar{y}$. The same reasoning as used in deriving formula (9) leads us to the formula

$$(27) \qquad V(\bar{y}) = \frac{1}{m} \{S_b^2 (1-\frac{m}{M}) + \frac{1}{M} \sum_{i=1}^{M} (MW_i)^2 V(\bar{y}_i)\},$$

where S_b^2 and $V(\bar{y}_i)$ are given by

$$(28) \qquad S_b^2 = \frac{1}{M-1} \sum_{i=1}^{M} (MW_i \bar{y}_i - \bar{Y})^2,$$

$$(29) \qquad V(\bar{y}_i) = \sum_{j=1}^{n_i} (w_{ij}/W_i)^2 V(\bar{y}_{ij}).$$

Here, $V(\bar{y}_{ij})$ is the variance of $\bar{y}_{ij}$ and we neglected the finite population correction. (The finite population corrections go to zero as the population size increases.)

The above formulas are not expressed in terms of quantities being calculable from a sample. In order to estimate the error from a sample, we need estimators of S_b^2, $V(\bar{y}_i)$, and $V(\bar{y})$. Let n_{ij} denote the number of elements sampled from the j-th cell in the i-th cluster. It is clear that, unless $m > 1$ and $n_{ij} > 1$, we have no unbiased estimators of $V(\bar{y})$ and $V(\bar{y}_i)$. If $m > 1$, $n_{ij} > 1$, and elements are sampled randomly in each cell, we have

$$(30)\qquad v(\bar{y}) = \frac{1}{m}\{s_b^2\,(1-\frac{m}{M}) + \frac{1}{m}\sum_{i=1}^{m}(MW_i)^2 v(\bar{y}_i)\},$$

$$(31)\qquad s_b^2 = \frac{1}{m-1}\sum_{i=1}^{m}(MW_i\bar{y}_i - \bar{y})^2,$$

$$(32)\qquad v(\bar{y}_i) = \sum_{j=1}^{n_i}(w_{ij}/W_i)^2 s_{ij}^2/n_{ij},$$

as estimators of $V(\bar{y})$, S_b^2 and $V(\bar{y}_i)$, respectively, where s_{ij}^2 is the estimator of the variance of y_{ij}.

Practically, however, it is quite often that $n_{ij} = 1$, i.e., only one element is sampled from a cell, and it may even happen that $m = 1$, i.e., only one cluster is sampled. We, therefore, present practical (biased) estimators for the case of $n_{ij} = 1$.

If $p_1, p_2, \cdots, p_d$ are sufficiently large, we may neglect the intracell variance because the cell is sufficiently small, that is, we set s_{ij}^2 to zero. This approximation gives the following underestimating estimator:

$$(33)\qquad v_1(\bar{y}) = (s_b^2/m)(1-\frac{m}{M}).$$

This estimator can be used only when more than one cluster is sampled.

The second estimator is obtained by regarding a cluster as a random sample. The cells in a cluster are highly uniformly distributed over the parameter space, and the discussion in section 2.3 tells us that this approximation is not bad in most cases. Then, variances $V(\bar{y}_i)$ and $V(\bar{y})$ may be estimated by

$$(34)\qquad v_2(\bar{y}_i) = \frac{1}{n_i}\{\frac{1}{n_i-1}\sum_{j=1}^{n_i}(\frac{n_i w_{ij}}{W_i}y_{ij} - \bar{y}_i)^2\} \equiv \frac{s_i'^2}{n_i},$$

$$(35)\qquad v_2(\bar{y}) = \frac{1}{m^2}\sum_{i=1}^{m}(MW_i)^2\frac{s_i'^2}{n_i},$$

respectively. The estimator (35) can be used only when $m = 1$.

The third formula is obtained by combining formulas (33) and (35):

$$(36)\qquad v_3(\bar{y}) = \frac{1}{m}\{s_b^2\,(1-\frac{m}{M}) + \frac{1}{m}\sum_{i=1}^{m}(MW_i)^2\frac{s_i'^2}{n_i}\},$$

where $s_i'^2$ is defined in (34). This estimator can be used only when $m > 1$. Note that the variance of the sample maen in stratified sampling is always smaller than that in simple random sampling. That is, $V(\bar{y}_i) \leqq V_2(\bar{y}_i)$, where $V_2(\bar{y}_i)$ is the same as $v_2(\bar{y}_i)$ defined in (34) except that the averaging is done over all the elements contained in the i-th cluster. Therefore, formula (36) is an overestimating estimator of $V(\bar{y})$.

In applying formula (26), the author finds that it is often better to use the estimator

$$(37) \qquad \bar{y}' = \sum_{i=1}^{m} W_i \bar{y}_i \Big/ \Big(\sum_{i=1}^{m} W_i \Big)$$

than to use (26). The reason is as follows. Putting $y(x_1, x_2, \cdots, x_d)$ to 1 in (26), we have the theoretical identity

$$\sum_{i=1}^{m} MW_i = m.$$

This identity is, however, not always satisfied in actual cases, which often causes a considerable error in $\bar{y}$ if M is much greater than m. Using formula (37), the above identity is always satisfied.

§5. Practical aspects

5.1. Merits of the systematic sampling proposed

Our sampling method has various merits.

i) Applicability of various accuracy increasing techniques. In order to attain accuracy without increasing the sample size, it is generally necessary to utilize information about the population and design an effective sampling method. For example, in stratified sampling, a high accuracy will be obtained if we perform sampling in such a way that the number of elements sampled from each stratum is proportional to the square root of the intrastratum variance. Thus, to collect information about local behavior of the population is a key for using various accuracy increasing techniques. Such information is usually collected by

sampling and accumulated as one-dimensional or, at most, two-dimensional data [3], [4], [5], [6]. That is, the information in the d-dimensional space is projected on one- or two-dimensional spaces. If the information is collected by square-grid sampling, the data will be accumulated in several narrow regions in each axis. Such data do not give enough information about the local behavior of the population. On the other hand, the information is collected from the population highly uniformly in our method and projected over all strata w.r.t. each axis highly evenly.

ii) Predictability of sampling. Let $\hat{c} = C[\hat{j}_1,\hat{j}_2,\cdots,\hat{j}_d]$ be one of the leading cells and let $c = C[j_1,j_2,\cdots,j_d]$ be the cell to be sampled ℓth by algorithm SYSCLUST with the initial cell $\hat{c}$. We can calculate $(j_1,j_2,\cdots,j_d)$ by solving simple coupled equations. Let ℓ_i be the number of times of transfer from a stratum w.r.t. x_i to another in the course of sampling from $\hat{c}$ to c. Then, referring to the discussion preceding the lemma 2, we have the following relations:

$$\ell_d = \ell,$$

$$\ell_i k_i = \ell_{i-1} p_i + r_i, \quad 0 \leqq r_i < p_i, \ i=d,d-1,\cdots,1, \quad 0 \leqq \ell_0 < p_1.$$

From these relations, we can successively calculate ℓ_i and r_i, i=d,d-1, $\cdots$,1. The set $(r_d,r_{d-1},\cdots,r_1)$ gives $(j_d,j_{d-1},\cdots,j_1)$ directly, for example, $j_i = \hat{j}_i + r_i$ if $\hat{j}_i + r_i \leqq p_i$. That is, any cell to be sampled is predictable. Similarly, for a given cell $c = C[j_1,j_2,\cdots,j_d]$, we can predict whether c is sampled or not. Eliminating ℓ_d, ℓ_{d-1}, $\cdots$, and ℓ_1 in the above relations, we have

$$\ell(k_d k_{d-1}\cdots k_1) = r_d(k_{d-1}k_{d-2}\cdots k_1) + (p_d)r_{d-1}(k_{d-2}\cdots k_1)$$

$$+ (p_d p_{d-1})r_{d-2}(k_{d-3}\cdots k_1) + \cdots$$

$$+ (p_d p_{d-1}\cdots p_2)r_1 + (p_d p_{d-1}\cdots p_1)\ell_0$$

as a necessary condition for ℓ, r_i, i=d,d-1,$\cdots$,1, and ℓ_0. Conversely, if ℓ, ℓ_0, $0 \leqq \ell_0 < p_1$, and a set $(r_d,r_{d-1},\cdots,r_1)$, $0 \leqq r_i < p_i$,

$i=d,d-1,\cdots,1$, satisfying the above condition are given, we can calculate ℓ_i, $i=d,d-1,\cdots,1$, uniquely. Therefore, by converting $(j_1,j_2,\cdots,j_d)$ to $(r_1,r_2,\cdots,r_d)$ and by checking the above condition, we can predict whether c is sampled or not. The predictability of sampling makes our method quite flexible in actual applications.

iii) Applicability of parallel sampling. In our method, all the N cells are divided into M nonoverlapping clusters and each cluster is specified by one of the leading cells. Therefore, if only we choose m different leading cells initially, we can sample m clusters parallelly.

iv) Controllability of sample size. Controllability of the sample size is directly related to controllability of the computing time, hence it is important in high dimensions where large sample sizes are often necessary. Our experience shows that the size of a cluster in our method can be controlled within the accuracy of about 0.1. On the other hand, in square-grid sampling for example, the size of a sample is m^d with m an integer and it changes largely when m is changed.

v) Sequential access to the memory space. Suppose the population $\mathfrak{Y} = \{y(x_1,x_2,\cdots,x_d)\}$ is composed of a large amount of data. Then, in the scheme of random sampling, random access to a large data space is necessary, which is computationally quite inefficient. In our method, we can make an ordering of the data so that the access to the memory may be done sequentially. Another merit of scanning the parameter space sequentially is found in [12].

vi) When $y(x_1,x_2,\cdots,x_d)$ is a rational function in x_1, x_2, $\cdots$, and x_d, we can sample y quite fast so far as the sampling is done from the elements located at the centers of the cells. Let us explain this by the following simple example in three dimension:

$$y(x_1,x_2,\cdots,x_d) = \frac{1}{x_3^2 + 2x_3(x_2 + x_1) + x_2^2 + 2x_2x_1 + x_1^2 + 1}.$$

In our method, the value of x_3 is changed each time an element is sampled, while the value of x_2 remains unchanged for several changes of x_3, and the value of x_1 changes only p_1 times. Therefore, the following FORTRAN program will evaluate $y(x_1,x_2,\cdots,x_d)$ efficiently.

```
      FUNCTION Y(I)
C     "I" IS THE SMALLEST OF INDICES OF XI WHICH ARE
C         CHANGED BY PROCEDURE "NEXTCELL".
      COMMON X1,X2,X3,A,B,C,D
      GO TO (1,2,3),I
    1 A = X1*X1 + 1
      B = 2*X1
    2 C = X2*(X2+B) + A
      D = 2*(X2+X1)
    3 Y = 1/(X3*(X3+D) + C)
      RETURN
      END
```

5.2. Numerical investigations

We have tested our method by the following three functions:

$$\text{(A)} \qquad y(x_1,x_2,\cdots,x_d) = \prod_{i=1}^{d}(2x_i),$$

$$\text{(B)} \qquad y(x_1,x_2,\cdots,x_d) = \prod_{i=1}^{d}\{1 + b(1-6x_i+6x_i^2)\}, \quad b = 1,$$

$$\text{(C)} \qquad y(x_1,x_2,\cdots,x_d) = \prod_{i=1}^{d}\{1 + c(x_i-3x_i^2+2x_i^3)\}, \quad c = 8,$$

where the population mean is 1 for all examples and the population variance is $(4/3)^d-1$ for A, $(1+b^2/5)^d-1$ for B, and $(1+c^2/210)^d-1$ for C, respectively. We set $p_1 = p_2 = \cdots = p_d = p$, and determine the increments $k_2, k_3, \cdots, k_d$ so that $k_2 \leqq k_3 \leqq \cdots \leqq k_d \leqq (k_2+1)$ and the average cluster size may be slightly greater than n_c.

Tables I and II show results of sampling in eight dimension.

Table I		Table II

We see that, for A and C, our method gives the results which are not so much different from those expected from simple random sampling. While, the sample means for B are systematically smaller than the population mean. This is due to that function B has narrow and high peaks at $x_i = 0$ and 1, $i=1,2,\cdots,d$. Hence, the same trend will be observed if we apply simple random sampling to B.

Using functions A, B and C, and setting p to 23, 31, 43, d to

4, 6, 8, 10, m to 2, 4, and n_c to 10000, 20000, 40000, 80000, we have tested our method many times. The results show almost the same trends as those observed in tables I and II. We, however, noticed the following two phenomena: the sample means for A and C obtained by our method are often much more accurate than those expected in simple random sampling; for each example, the sample means are sometimes deviated much from the population mean, while the estimated values of the errors are not deviated much. We observed that the latter phenomenom is related to the selection of initial cells. The set of leading cells $\{C[1,j_2,j_3,\cdots,j_d] \mid 1 \leqq j_i \leqq k_i,\ i=2,3,\cdots,d\}$ form a hyper-rectangle. If a cell located closely to a corner of this hyper-rectangle is selected as the initial cell, then the sample mean is often deviated much. This phenomenom is an intrisic fault of the systematic sampling. In actual applications, therefore, we had better choose initial cells carefully. The author recommends to sample the initial cells not randomly but uniformly from the set of leading cells.

References

[1] W.G. Cochran, Sampling techniques, 3rd edition, John Wiley & Sons, New York, 1977.

[2] M.N. Murthy, Sampling theory and methods, Statistical Publishing Society, Calcutta, India, 1967.

[3] J.M. Hammersley and D.C. Handscomb, Monte Carlo methods, John Wiley, New York, 1964.

[4] B.E. Lautrup, "An adaptive multidimensional integration technique," Proceedings of the 2nd Colloquium on Advanced Computing Methods in Theoretical Physics, Marseille, France, 1971.

[5] T. Sasaki, "Multidimensional Monte Carlo integration based on factorized approximation functions," SIAM J. Numer. Anal. 15 (1978), pp.938-952.

[6] G.P. Lepage, "A new algorithm for adaptive multidimensional integration," J. Comp. Phys. 27 (1978), pp.192-203.

[7] P.G. Homeyer and C.A. Black, "Sampling replicated field experiments on oats for yield determinations," Proc. Soil Sci. Soc. America 11 (1946), pp.341-344.

[8] F. Yates, Sampling methods for censuses and surveys, 3rd edition, Charles Griffin and Co., London, 1960.

[9] H.D. Patterson, "The errors of lattice sampling," J. Roy. Stat. Soc. B16 (1954), pp.140-149.

[10] J.F. Koksma, "A general theorem from the theory of uniform distribution modulo 1," Mathematica Zutphen B, 11 (1949), pp.7-11.

[11] E. Hlawka, "Funktionen von beschränkter Variation in der Theorie der Gleichverteilung," Ann. Mat. Pura Appl. (4), 54 (1961), pp.325-334; see, also
A.H. Stroud, Approximate calculation of multiple integrals, Prentice-Hall Inc., Englewood Cliffs, New Jersey, 1971, Ch.6.

[12] T. Sasaki, "Multidimensional integration of functions with narrow and high peaks," preprint of IPCR (in preparation).

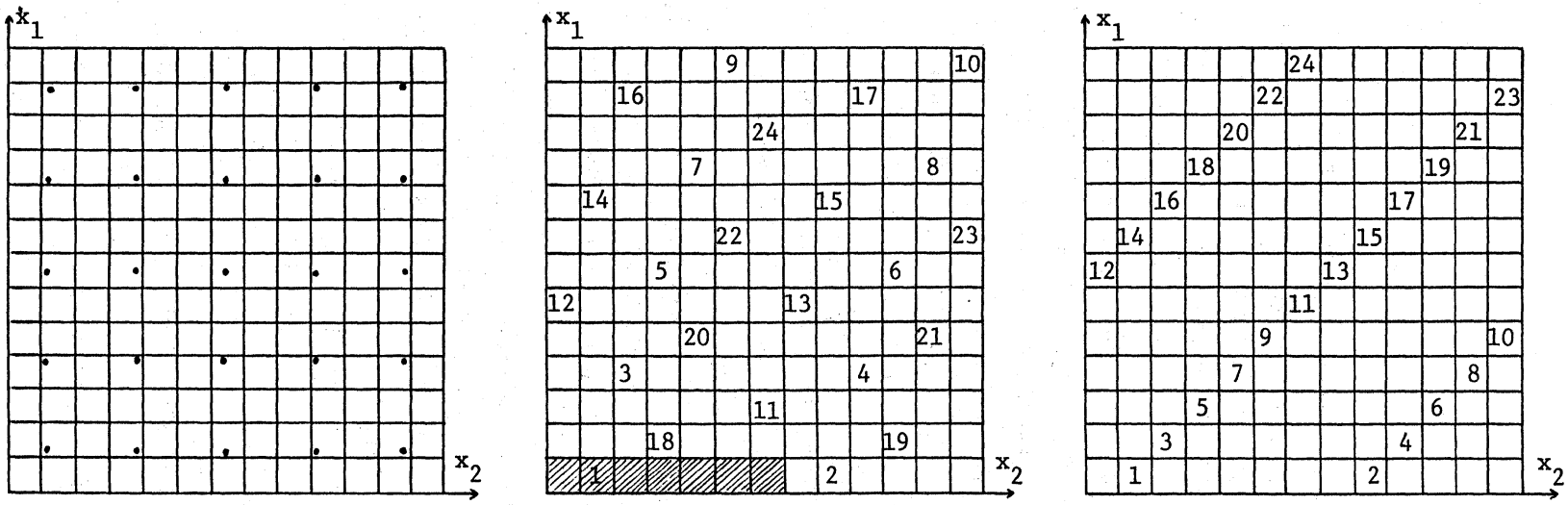

Fig.1. Illustration of uniformity of the distribution of sampled elements: a) elements at cross points of a square grid are sampled, b) elements at the center of the cells numbered are sampled in this order, c) sampling is same as b except that the increment in the x_1 direction is 1.

Problem A

cluster 1 :	$\bar{y}_1 = 0.9897$,	$\sqrt{v_2(\bar{y}_1)} = 0.0271$	(0.0298)
cluster 2 :	$\bar{y}_2 = 1.019$,	$\sqrt{v_2(\bar{y}_2)} = 0.0288$	(0.0298)
total :	$\bar{y}' = 1.0046$,	$\sqrt{v_3(\bar{y}')} = 0.0234$	(0.0211)

Problem B

cluster 1 :	$\bar{y}_1 = 0.9757$,	$\sqrt{v_2(\bar{y}_1)} = 0.0161$	(0.0181)
cluster 2 :	$\bar{y}_2 = 0.9757$,	$\sqrt{v_2(\bar{y}_2)} = 0.0170$	(0.0180)
total :	$\bar{y}' = 0.9757$,	$\sqrt{v_3(\bar{y}')} = 0.0118$	(0.0128)

Problem C

cluster 1 :	$\bar{y}_1 = 0.9922$,	$\sqrt{v_2(\bar{y}_1)} = 0.0268$	(0.0271)
cluster 2 :	$\bar{y}_2 = 1.015$,	$\sqrt{v_2(\bar{y}_2)} = 0.0285$	(0.0270)
total :	$\bar{y}' = 1.0036$,	$\sqrt{v_3(\bar{y}')} = 0.0219$	(0.0191)

Table I. Results of sampling in eight dimension (d = 8). Other parameters are p = 31 (the number of intervals in each axis), m = 2 (the number of clusters sampled), and n_c = 10000 (the expected size of a cluster). The population mean is 1. The estimators (25) and (34) are used for each cluster, and (37) and (36) are used for the cluster sum. For comparison, the expected error in simple random sampling are appended in parentheses. The cluster size is 10095 for cluster 1 and 10150 for cluster 2.

Problem A

cluster 1 :	$\bar{y}_1 = 0.9497$,	$\sqrt{v_2(\bar{y}_1)} = 0.0266$	(0.0298)
cluster 2 :	$\bar{y}_2 = 1.025$,	$\sqrt{v_2(\bar{y}_2)} = 0.0298$	(0.0298)
cluster 3 :	$\bar{y}_3 = 1.005$,	$\sqrt{v_2(\bar{y}_3)} = 0.0288$	(0.0298)
cluster 4 :	$\bar{y}_4 = 0.9773$,	$\sqrt{v_2(\bar{y}_4)} = 0.0278$	(0.0298)
total :	$\bar{y}' = 0.9893$,	$\sqrt{v_3(\bar{y}')} = 0.0166$	(0.0149)

Problem B

cluster 1 :	$\bar{y}_1 = 0.9642$,	$\sqrt{v_2(\bar{y}_1)} = 0.0158$	(0.0181)
cluster 2 :	$\bar{y}_2 = 0.9295$,	$\sqrt{v_2(\bar{y}_2)} = 0.0154$	(0.0180)
cluster 3 :	$\bar{y}_3 = 0.9540$,	$\sqrt{v_2(\bar{y}_3)} = 0.0154$	(0.0181)
cluster 4 :	$\bar{y}_4 = 0.9688$,	$\sqrt{v_2(\bar{y}_4)} = 0.0166$	(0.0181)
total :	$\bar{y}' = 0.9541$,	$\sqrt{v_3(\bar{y}')} = 0.0088$	(0.0090)

Problem C

cluster 1 :	$\bar{y}_1 = 1.006$,	$\sqrt{v_2(\bar{y}_1)} = 0.0283$	(0.0271)
cluster 2 :	$\bar{y}_2 = 0.9944$,	$\sqrt{v_2(\bar{y}_2)} = 0.0290$	(0.0270)
cluster 3 :	$\bar{y}_3 = 0.9882$,	$\sqrt{v_2(\bar{y}_3)} = 0.0283$	(0.0271)
cluster 4 :	$\bar{y}_4 = 0.9862$,	$\sqrt{v_2(\bar{y}_4)} = 0.0276$	(0.0271)
total :	$\bar{y}' = 0.9938$,	$\sqrt{v_3(\bar{y}')} = 0.0144$	(0.0135)

Table II. Results of sampling in eight dimension. Parameters are the same as in Table I except that m = 4. The cluster size is 10095 for cluster 1, 10151 for cluster 2, 10097 for cluster 3, and 10095 for cluster 4.