

## Optimal Design of the Adaptive Searching Estimation in Spatial Sampling

Pyong Namkung<sup>1)</sup> and Jong Seok Byun<sup>2)</sup>

### Abstract

The spatial population existing in a plane area, such as an animal or aerial population, have certain relationships among regions which are located within a fixed distance from one selected region. We consider with the adaptive searching estimation in spatial sampling for a spatial population. The adaptive searching estimation depends on values of sample points during the survey and on the nature of the surfaces under investigation. In this paper we study the estimation by the adaptive searching in a spatial sampling for the purpose of estimating the area possessing a particular characteristic in a spatial population. From the viewpoint of adaptive searching, we empirically compare systematic sampling with stratified sampling in spatial sampling through the simulation data.

*Keywords* : Spatial sampling, Adaptive searching estimation, Subrectangles, Adaptive searching observation, Systematic sampling, Stratified random sampling.

### 1. Introduction

The basic purpose of sample survey is to estimate some characteristics of a population by observing only part of the population. We consider the estimation of the area possessing a particular characteristic in a spatial population. The spatial population existing in a plane area, like an animal or aerial population, have certain relationships among regions which are located within a fixed distance from one selected region. The neighborhoods of a sample point are related because the spatial population has spatial correlation among sample points within population.

---

1) Professor, Department of Statistics, Sungkyunkwan University, Seoul, 110-745, Korea.  
E-mail : namkung@yurim.skku.ac.kr

2) Associate Professor, Department of Statistics and Information, Hanshin University, Kyunggi-do, 447-791, Korea.  
E-mail : jsbyun@hucc.hanshin.ac.kr

The spatial sampling is to select sample points that is representative of a coordinate as sampling units in a plane area. Quenouille (1949), Bellhouse (1977, 1981) and Koop (1990) described the sampling methods of sample points in a two-dimensional space. They proposed that unrestricted conventional sampling method may be employed in either direction. Furthermore, the sampling points in either direction may either be aligned with one another or be unaligned, i.e., independently determined. They were concerned with estimation by the observation of the all sample points. But the adaptive searching is to observe the neighboring sample points when the current sample point satisfies the condition of an adaptive searching principle (Byun and Namkung, 1996). In adaptive searching, the observation of the sample points may depend on values of the variables of interest which are observed during the survey.

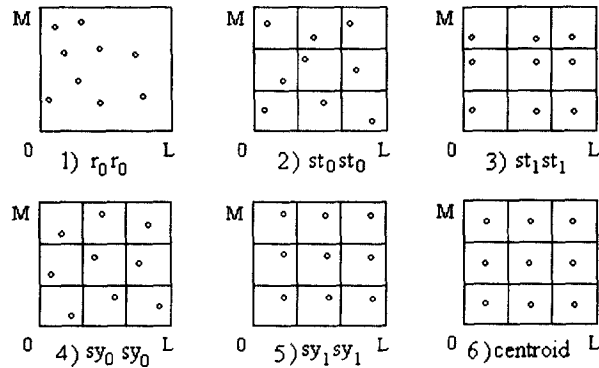
In this paper we shall be concerned with the estimation by the adaptive searching in a spatial sampling for the purpose of estimating the area possessing a particular characteristic in a spatial population. From the viewpoint of adaptive searching we will empirically compare systematic sampling with stratified sampling in spatial sampling through the simulation data.

## 2. Spatial Sampling

Spatial sampling is the method of the sampling of points in a plane area for the purpose of estimating the proportion of the area possessing a particular characteristic such as pollutions, crop acreages. Quenouille (1949) described briefly methods of sampling a plane area. He recognizes that unrestricted random (r), stratified random (st), and systematic sampling (sy) may be employed in either direction. The sampling points (units) in either direction may either be aligned with one another (the suffix 1), or be unaligned (the suffix 0), i.e., independently. Koop (1990) enumerates 21 different methods of sampling a plane including 12 methods considered by Quenouille. Koop (1990) and Quenouille (1949) described the methods for sampling points in a rectangular plane.

Because we shall select the sample points distributed over the plane area, we consider 4 methods,  $st_0st_0$ ,  $st_1st_1$ ,  $sy_0sy_0$ ,  $sy_1sy_1$ , and centroid sampling of points in the estimation of area.

On a two-dimensional surface there is a rectangular area of length  $L$  and breath  $M$  containing an elementary closed curve  $\alpha$  of area  $A_\alpha$  which is irregular (see Fig. 3.2). The measurements  $L$  and  $M$  is known but  $A_\alpha$  is unknown. We are interested in estimating the area of  $\alpha$  through a sample of points taken within this rectangular area, measuring  $A=LM$ , according to the methods explained in Fig. 2.1.



< Fig. 2.1 > 6 Samples of 9 points in a Rectangular Plane

**2.1 Fully Random (  $r_0 r_0$  )**

The theory for this method is very elementary. A pair of random coordinates  $(x, y)$ , for  $0 \leq x \leq L, 0 \leq y \leq M$ , determines a random point. Let  $n$  points be independently located in rectangular by  $n$  independent pairs of random coordinates, and suppose  $g$  points fall inside  $\alpha$ . Then an unbiased estimate of  $A_\alpha$  is

$$\hat{A}(r_0 r_0) = Ag/n, \tag{2.1}$$

since  $E(g/n) = p = A_\alpha/A$ , and its variance is

$$Var(r_0 r_0) = A^2 p(1-p)/n. \tag{2.2}$$

Note that  $g$  follows binomial distribution.

**2.2 Stratified Random Sampling in Both Directions (  $st_0 st_0$  )**

The rectangular area  $A$  is now divided into  $n=ab$  congruent subrectangles with  $a$  rows of subrectangles parallel to the  $X$ -axis on  $OL$  and  $b$  columns of subrectangles parallel to the  $Y$ -axis on  $OM$ . Clearly this division constitutes stratification in both perpendicular directions. Within each of the resulting  $n$  rectangles (or strata), one point is located at random independently of all other  $n-1$  points.

Let  $(ij)$  be the subrectangles of the  $i$ th row and  $j$ th column ( $i = 1, 2, \dots, a, j = 1, 2, \dots, b$ ), and let  $(ij)$  be the random point located in stratum  $(ij)$ . We also let  $\alpha_{ij}$  be the subregion of

$\alpha$  that is found within stratum  $(ij)$  as a result of the foregoing division of  $A$ . In stratum  $(ij)$  let  $z_{ij}$  be a characteristic random variable such that

$$z_{ij} = \begin{cases} 1, & \text{if point } (ij) \in \alpha_{ij}, \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore let  $\Pr\{z_{ij} = 1\} = p_{ij}$ . It is easy to show that

$$p = \frac{1}{n} \sum_{i=1}^a \sum_{j=1}^b p_{ij}, \quad (2.3)$$

$$E(z_{ij}) = p_{ij}, \text{ and } \text{Var}(z_{ij}) = p_{ij}(1 - p_{ij}) \text{ for all } (ij).$$

An unbiased estimate of  $A_\alpha$  is

$$\widehat{A}(st_0st_0) = \frac{A}{n} \sum_{i=1}^a \sum_{j=1}^b z_{ij}, \quad (2.4)$$

with variance

$$\text{Var}(st_0st_0) = \frac{A^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b \text{Var}(z_{ij}) = \frac{A^2}{n^2} \sum_{i=1}^a \sum_{j=1}^b p_{ij}(1 - p_{ij}). \quad (2.5)$$

### 2.3 Stratified Random Sampling with Alignments in Both Directions ( $st_1st_1$ )

The  $ab$  points are located by first taking  $b$  independent  $x$ -coordinates on  $OL$ , each corresponding to the  $b$  subrectangles on the  $X$ -axis, and  $a$  independent  $y$ -coordinates on  $OM$ , each corresponding to the  $a$  subrectangles on the  $Y$ -axis. Consider  $ab$  points that are intersections of the  $b$  lines drawn parallel to the  $Y$ -axis with the  $a$  lines drawn parallel to the  $X$ -axis through each of the selected random points on the  $X$  and  $Y$ -axis respectively. Now select a random point in each of the  $ab$  strata.

Then an unbiased estimate of  $A_\alpha$  is

$$\widehat{A}(st_1st_1) = \frac{A}{n} \sum_{i=1}^a \sum_{j=1}^b z_{ij}, \quad (2.6)$$

with variance

$$\begin{aligned} \text{Var}(st_1 st_1) = \frac{A^2}{n^2} & \left( \sum_{i=1}^a \left\{ \sum_{j=1}^b \text{Var}(z_{ij}) + \sum_{j \neq i}^b \text{Cov}(z_{ij}, z_{ij}) \right\} \right. \\ & \left. + \sum_{i \neq j}^a \left\{ \sum_{j=1}^b \text{Cov}(z_{ij}, z_{ij}) + \sum_{j \neq i}^b \text{Cov}(z_{ij}, z_{ij}) \right\} \right). \end{aligned} \quad (2.7)$$

Because points between parallel strata are independently chosen, there is a covariance between the  $b$  points in each of the  $a$  sets of strata that are parallel to the  $X$ -axis. Additionally, there is a covariance between the  $a$  points in each of the  $b$  sets of strata that are parallel to the  $Y$ -axis.

#### 2.4 Systematic Sampling without Alignments in Both Directions ( $sy_0 sy_0$ )

As in  $st_0 st_0$  of section 2.2, the division of the rectangular area into  $a = ab$  equal subrectangles results in a sets of subrectangles parallel to the  $X$ -axis and  $b$  sets parallel to the  $Y$ -axis. We note again that the dimensions of each subrectangle are  $L/b$  in the direction of the  $X$ -axis and  $M/a$  in the direction of the  $Y$ -axis. To locate  $ab$  points systematically in either perpendicular directions without alignment, we first choose randomly and independently with equal probability two sets of  $x$ - and  $y$ -coordinates, and those are

$$\{x_i: 0 \leq x_i \leq L/b, i = 1, 2, \dots, a\}$$

and

$$\{y_j: 0 \leq y_j \leq M/a, j = 1, 2, \dots, b\}.$$

Then the unaligned systematic sample of  $ab$  points is given by the set of  $ab$  possible coordinate pairs

$$\left( x_i + \frac{(j-1)L}{b}, y_j + \frac{(i-1)M}{a} \right), (i = 1, 2, \dots, a; j = 1, 2, \dots, b).$$

We see that points are systematically located in either perpendicular direction but without alignment because of an irregularly varying  $y$ - or  $x$ -coordinates in each respective direction. The  $b$  points in each set of subrectangles parallel to the  $X$ -axis are not independent of each other because of having a constant  $y$ -coordinate and similarly for points in subrectangles parallel to the  $Y$ -axis. But points belonging to different rows and columns are independent.

An unbiased estimate of  $A_\alpha$  is

$$\hat{A}(sy_0 sy_0) = \frac{A}{n} \sum_{i=1}^a \sum_{j=1}^b z_{ij}, \quad (2.8)$$

with variance

$$\begin{aligned} \text{Var}(sy_0 sy_0) = & \frac{A^2}{n^2} \left( \sum_{i=1}^a \left\{ \sum_{j=1}^b \text{Var}(z_{ij}) + \sum_{j \neq i}^b \text{Cov}(z_{ij}, z_{ij'}) \right\} \right. \\ & \left. + \sum_{i \neq i'}^a \sum_{j=1}^b \text{Cov}(z_{ij}, z_{i'j}) \right). \end{aligned} \quad (2.9)$$

Although the covariance in (2.7) looks like the same as those in (2.9), they are essentially different because  $\text{Cov}(z_{ij}, z_{ij'})$  in (2.7) are dependent on  $X$ -axis and  $\text{Cov}(z_{ij}, z_{ij'})$  in (2.9) are dependent on  $Y$ -axis.

## 2.5 Systematic Sampling with Alignment in Both Directions ( $sy_1 sy_1$ )

The  $ab$  points are determined by first selecting a pair of random coordinates  $(x,y)$  such that  $0 \leq x \leq L/b$  and  $0 \leq y \leq M/a$ . Then these points are determined by the  $ab$  coordinate pairs

$$\left( x + \frac{(j-1)L}{b}, y + \frac{(i-1)M}{a} \right), \quad (i = 1, 2, \dots, a; j = 1, 2, \dots, b).$$

For any given subrectangles the point is determined fully at random. However, with respect to whole area the  $ab$  points are not mutually independent.

An unbiased estimate of  $A_a$  is

$$\hat{A}(sy_1 sy_1) = \frac{A}{n} \sum_{i=1}^a \sum_{j=1}^b z_{ij}, \quad (2.10)$$

with variance

$$\begin{aligned} \text{Var}(sy_1 sy_1) = & \frac{A^2}{n^2} \left( \sum_{i=1}^a \left\{ \sum_{j=1}^b \text{Var}(z_{ij}) + \sum_{j \neq j'}^b \text{Cov}(z_{ij}, z_{ij'}) \right\} \right. \\ & \left. + \sum_{i \neq i'}^a \left\{ \sum_{j=1}^b \text{Cov}(z_{ij}, z_{i'j}) + \sum_{j \neq j'}^b \text{Cov}(z_{ij}, z_{i'j'}) \right\} \right). \end{aligned} \quad (2.11)$$

## 2.6 Centroid (Systematic) Sampling

Centroid sampling is a special case of systematic sampling with alignment in both directions. As in  $sy_0 sy_0$  and  $sy_1 sy_1$  of sections 2.4 and 2.5, common form of systematic sampling is to partition  $A$  into square strata, then the  $ab$  points are determined one point per strata in centre of each square. An unbiased estimator of  $A_a$  and its variance are the same as those of  $\hat{A}(sy_1 sy_1)$ .

### 3. Adaptive Searching Estimation

We consider the point-counting technique by selecting one sample point per partitioned cell to estimate the interesting area in a spatial population. Let  $z_{ij}$  be a characteristic random variable such that for each  $(ij)$

$$z_{ij} = \begin{cases} 1, & \text{if point } (ij) \text{ belongs to the surveyed region,} \\ 0, & \text{otherwise.} \end{cases}$$

The value of observed sample point per each cell is 1 or 0. Therefore sample point in each cell follows Bernoulli distribution.

#### 3.1 Adaptive Searching Observation Procedure

If a spatial population is partitioned into  $n$  cells, we assume that the probability of an isolated area is 0 and Figure 3.1 explains the isolated area.

$$\begin{array}{ccc} & 0 & 1 \\ 0 & 1 & 0 \\ & 0 & 1 \end{array}$$

<Fig. 3.1> Property of an isolated area

We determine coordinates of sample points by spatial sampling with respect to initial cells sampled from partitioned spatial population, and then observe the sample value of those points. The observed sample points are either 1 or 0. All observed values for the initial selected cells are 1, or some are 1 and some are 0.

In adaptive searching estimation, the adaptive searching condition is determined from the values( $z$ ) of the observed sample points as follow:

$$z = \begin{cases} 0, & \text{if the observation is started from external region,} \\ 1, & \text{otherwise.} \end{cases}$$

We continue to search neighboring cells as long as the observed values of the cells satisfy the adaptive searching condition. The principle for adaptive searching observation follows:

#### <Principles for Adaptive Searching Observation>

First, initial cells are selected by simple random sampling. But for simplicity an initial cells are limited to the four edge corner cells. A sample point per the initial cell is selected by





consisting of observed cells ( $\psi_m$ ) and unobserved cells ( $\psi_{n-m}$ ). The unobserved cells can be predicted from observing condition by adaptive searching. The adaptive searching estimator is

$$\begin{aligned}\widehat{A}^a &= A_m \cdot \frac{1}{m} \sum_{(ij) \in \psi_m} z_{ij} + A_{n-m} \cdot \frac{1}{n-m} \sum_{(ij) \in \psi_{n-m}} z_{ij} \\ &= \widehat{A}_m^a + \widehat{A}_{n-m}^a.\end{aligned}\quad (3.1)$$

Under the assumption that the spatial population can be partitioned into cells with the same shape and size, it is easy to show that the adaptive searching estimator  $\widehat{A}^a$  is an unbiased estimator of  $A^a$ . The variance of the adaptive searching estimator  $\widehat{A}^a$  is defined as the sum of the variance of observed and unobserved network. But the value of the unobserved network is constant and the variance of the adaptive searching estimator  $\widehat{A}^a$  is defined as follows

$$\begin{aligned}Var(\widehat{A}^a) &= A_m^2 \cdot \frac{1}{m^2} \left[ \sum_{(ij) \in \psi_m} Var(z_{ij}) \right] \\ &= A_m^2 \cdot \frac{1}{m^2} \left[ \sum_{(ij) \in \psi_m} p_{ij} (1 - p_{ij}) \right].\end{aligned}\quad (3.2)$$

Suppose two more samples of points are independently selected, and if  $\widehat{A}_1^a, \widehat{A}_2^a, \dots, \widehat{A}_K^a$  are unbiased estimates of  $A^a$ , then an unbiased estimate of  $Var(\widehat{A}^a)$  is

$$\widehat{Var}(\widehat{A}^a) = \frac{1}{K \cdot (K-1)} \left[ \sum_{k=1}^K (\widehat{A}_k^a - \overline{A}^a)^2 \right], \quad (3.3)$$

where  $\overline{A}^a = \frac{1}{K} \sum \widehat{A}_k^a$ .

By Byun and Namkung (1996), the number of adaptive searching observation is reduced and the result of adaptive searching estimation in a spatial population is the same as that of the non-adaptive method.

#### 4. Empirical Study

To compare the efficiency of the adaptive searching estimation in spatial sampling, we estimate a contaminant area in the spatial population. We consider four spatial populations in Table 4.1.

Table 4.1 Type of the spatial population in simulation data

Shape of Population	Spatial Population			Sample design	
	Area	Contaminant Area	P	Sample size	Partitioned cell Area
Convex	30 × 30	477	0.5300	100	3 × 3
	50 × 50	477	0.1908	100	5 × 5
Concave	30 × 30	432	0.4800	100	3 × 3
	50 × 50	432	0.1728	100	5 × 5

Each of the sample size is 100. Therefore we split the spatial population into  $10 \times 10$  squares and then estimate the contaminant area by using the adaptive searching method in that area. To select sample point per the partitioned cell, we consider 4 spatial sampling methods,  $st_0st_0$ ,  $st_1st_1$ ,  $sy_0sy_0$ ,  $sy_1sy_1$ , and centroid systematic sampling. From the viewpoint of adaptive searching estimation we compare systematic sampling with stratified sampling. Each simulation is repeated 1,000 times for the computation of MSE and the values are mean of the repetition. Table 4.2 shows the result of the adaptive searching estimation for the contaminant area.

Table 4.2 The results of adaptive searching estimation for simulation data

Sample design			Stratified Sampling		Systematic Sampling		
			Unalignment	Alignment	Unalignment	Alignment	Centroid
C o n v e x	30 × 30	Obs.	67.59	67.92	67.31	68.42	76.00
		Est.	503.09	498.76	503.66	496.36	441.00
		MSE	1692.98	2173.23	1423.48	1629.48	1296.00
	50 × 50	Obs.	94.07	94.16	93.91	93.68	90.00
		Est.	479.63	478.40	480.23	478.38	525.00
		MSE	2229.88	4454.15	1368.73	866.13	2304.00
C o n c a v e	30 × 30	Obs.	73.60	75.50	73.07	76.81	78.00
		Est.	468.01	458.85	471.73	455.57	423.00
		MSE	3137.37	3004.05	2736.50	1417.42	81.00
	50 × 50	Obs.	95.18	95.54	95.32	95.14	91.00
		Est.	453.53	444.15	448.80	447.18	500.00
		MSE	3230.28	5463.40	2358.30	2717.93	4624.00

## 5. Conclusion and Comments

We studied the adaptive searching estimation in spatial sampling for a spatial population. The adaptive searching estimation depends on values of sample points during the survey and

the nature of the surfaces under investigation. As in the empirical study, the adaptive searching estimation of systematic sampling is more efficient than that of stratified sampling and if this method is applied to a convex population, then its the efficiency to the adaptive searching estimation increases. Despite the lack of firm theoretical results on the adaptive searching estimation, this method of spatial sampling in plane area are practical and potentially useful.

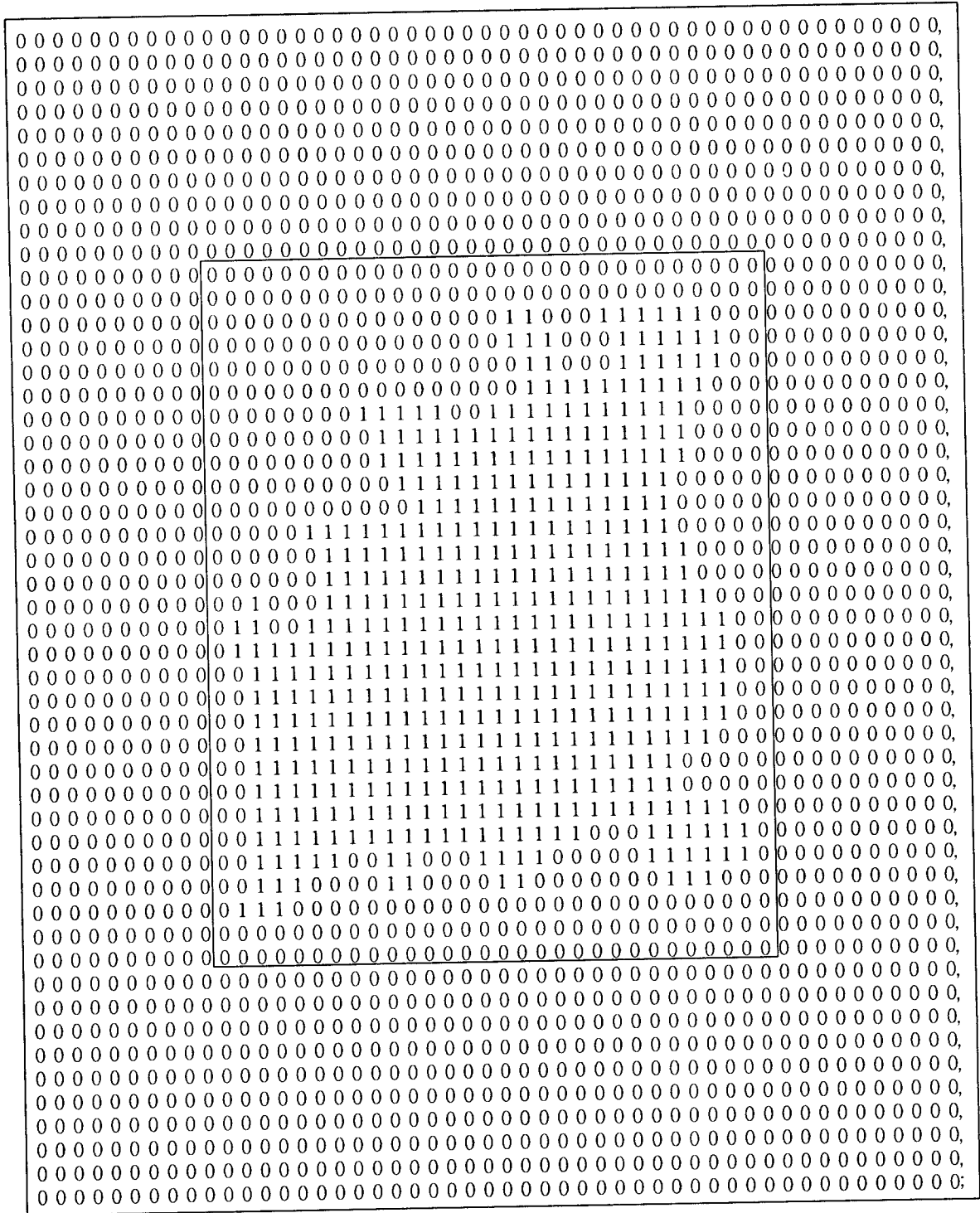
The theory considered in this paper can be applied to aerial and satellite photographs, and the extension of the present theory for a plane to three dimensional surfaces will give practical results.

### References

- [1] Bellhouse, D. R. (1977). "Some optimal designs for sampling in two dimensions," *Biometrika*, 64, 605-612.
- [2] Bellhouse, D. R. (1981). "Area estimation by point-counting techniques," *Biometrics*, 37, 303-312.
- [3] Byun, J. S. and Namkung, P. (1996). "Estimation using the method of adaptive searching observation," *The Korean Journal of Applied Statistics*, 9 (2), 145-159.
- [4] Koop, J. C. (1990). "Systematic sampling of two-dimensional surfaces and related problems," *Commun. Statist. - Theory and Methods*, 9 (5), 1701-1750.
- [5] Quenouille, M. H. (1949). "Problems in plane sampling," *Annals of Mathematical Statistics*, Vol. 20, 355-375.
- [6] Thompson, S. K. (1990). "Adaptive cluster sampling," *Journal of the American Statistical Association*, 85, 1050-1059.
- [7] Thompson, S. K. (1991a). "Adaptive cluster sampling : Designs with primary and secondary units," *Biometrics*, 47, 1103-1115.
- [8] Thompson, S. K. (1991b). "Stratified adaptive cluster sampling," *Biometrika*, 78, 389-397.
- [9] Thompson, S. K. (1992). *Sampling*, New York : Wiley.
- [10] Thompson, S. K., Ramsey, F. L. and Seber, G. A. F. (1992). "An adaptive procedure for sampling animal populations," *Biometrics*, 48, 1195-1199.
- [11] Thompson, S. K. and Seber, G. A. F. (1996). *Adaptive Sampling*, New York : Wiley.

< Appendix 1 ; Convex type >

- 1) Outer box ; 50 X 50
- 2) Inner box ; 30 X 30



< Appendix 2 ; Concave type >

- 1) Outer box ; 50 X 50
- 2) Inner box ; 30 X 30

