Adaptive Optimal Selection of Sampling Locations

Chang-Tai Chao, Steven K. Thompson, The Pennsylvania State University
Chang-Tai Chao, Department of Statistics, Pennsylvania State University
University Park, PA 16802 U.S.A. (chao@stat.psu.edu)

Key Words: Adaptive sampling, Model-based sampling, Optimal design, Sequential design, Log-normal distribution.

1. Introduction

Under a given population model, the optimal conventional sampling strategy would be to purposively select the sampling units to minimize the conditional mean square error given the sample s, $\mathrm{E}[(\hat{T}-T)^2|s]$. In the field of spatial statistics, the related problem is to optimally select n sample sites out of N possible sites to estimate or predict the population quantity of interest, for example, the total of population variable of interest (cf., Cressie 1993, p. 268-273, 319-322, Guttorp, et al. 1993, Sacks and Schiller 1988). With an isotropic spatial model and a spatial covariance function that is nonnegative and monotonically decreases as the distance increases, the optimal conventional sampling strategy will in general provide a systematic arrangement of sampling locations which spread throughout the study region as evenly as possible, with modifications to accommodate the boundary shape and other irregularities of the study region (Matérn 1986, Ripley 1976). However, it is possible to do still better with an adaptive design which takes the observed values into account as well, while the optimal conventional sampling strategy considers only the locations of sample sites(Zacks 1969, Basu 1969, Thompson and Seber 1996).

Zacks (1969) described the optimal sampling design, with a fixed sample size, under a population model with a given prior distribution. The optimal sampling design is in general a design which sequentially takes into account the observed values of the population variable of interest, that is, it is adaptive. Also Zacks (1969) gave a sufficient condition under which the optimal conventional (non-adaptive) design can be as good as the adaptive design. Necessary and sufficient condition for the optimal design to be non-adaptive with fixed sample size, along with an extension to a design in which the sample size is also adaptive, are given in Thompson (1988) and Thompson and Seber (1996). The essence of the optimal adaptive design is that at any point during the survey, one selects the remaining sampling sites by a procedure that minimize the mean square error (or other risk function) conditional on what has been observed so far. By doing so, the overall mean square error is minimized as well.

An example of a population model satisfying the necessary and sufficient condition for the optimal sampling strategy to be a conventional one is the multivariate normal distribution with known covariance function (Thompson and Seber 1996 p.242-5) and by extension the spatial Gaussian model with a known covariance or semivariance function. The reason is that the conditional covariance function of the unobserved values does not depend on the observed values. On the other hand, if the covariance structure is not known exactly but is specified by a prior distribution, then the conditional covariance function of the unobserved values will depend on the observed values (Cressie 1993 p.171). Such a model will in general no longer satisfy the necessary and sufficient condition for the optimal strategy to be non-adaptive. With many other models, including the log-normal or spatial log-Gaussian models, the optimal sampling is adaptive since the conditional variances and covariances of the unobserved values given the observed values do depend on the observed values.

We examine the optimal sampling strategy with an assumed spatial log-Gaussian model in the following sections. The log-Gaussian model is chosen for the illustrative purpose because of its practical importance in describing real phenomena, at least as an approximation, providing data having nonnegative values and skewed distributions. The population quantity of interest to be predicted is the total of the variable of interest. The relative efficiency of the optimal sampling strategy, which is adaptive, to the optimal conventional sampling strategy is computed for different covariance function parameters. In Section 2 the optimal sampling strategy, which is adaptive, is derived for the log-normal model. The optimal conventional strategy is also derived. Op-

Support for this research was provided by the National Science Foundation, grant DMS-9626102, and the National Institutes of Health, NIDA, grant RO1 DA09872.

timal sampling locations selected by these sampling designs are illustrated in Section 3; because of the nature of adaptive design, the sampling locations selected by the optimal adaptive design vary from realization to realization. The relative efficiency of the optimal strategy to the optimal conventional strategy is computed in Section 4. The relative efficiency, though always greater than one, is seen to depend on the "range of influence" spatial covariance function parameter. We also discuss some undergoing work of this optimal sampling strategy in Section 5.

2. Optimal Sampling Strategy with a Lognormal Model

For simplicity we assume that the population itself consists of a finite set of N units labeled $1, 2, \ldots N$. In the spatial setting this corresponds to partitioning the study region into N spatial units. Associated with the N units in the population is a vector of random variables $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$, a realization of which is denoted $\mathbf{y} = (y_1, y_2, \ldots, y_N)$. An assumed distribution $F(\mathbf{y})$ or probability density function $f(\mathbf{y})$ is referred to as the population model. A sample s of n units is selected from the N units in the population. The data $d = (s, \mathbf{y}_s)$ consist of the sample unit labels together with the associated values $\mathbf{y}_s$ of the variable of interest. For any possible sample s the design $p(s \mid \mathbf{y}_s)$ gives the probability that s is the sample selected. With an adaptive design this probability can depend on $\mathbf{y}_s$, the realized y-values observed in the sample. A conventional design does not depend on any y-values and can be denoted simply $p(s)$.

By sampling *strategy* for estimating or predicting a population quantity $T = T(\mathbf{Y})$ we mean a design p for selecting the sample together with an estimator or predictor $\hat{T} = \hat{T}(d)$. In particular, we consider prediction of the population total $T = \sum_{i=1}^{N} Y_i$ based on the observed values of the n units. An optimal strategy, for the purposes of this paper, will consist of a design and an unbiased estimator or predictor giving the smallest mean square error among the class of unbiased estimators or predictors and among a specified class of designs.

Consider an N-dimensional population vector

$$\mathbf{Y} = (Y_1, Y_2, \ldots\ldots, Y_N)'$$

under a lognormal spatial model with known mean vector and covariance matrix. The associated Gaussian random vector is

$$\begin{aligned} \mathbf{W} &= (W_1, W_2, \ldots\ldots, W_N)' \\ W_i &= \log Y_i, \qquad\qquad \forall i = 1, \ldots\ldots, N \end{aligned}$$

where

$$\mathbf{W} \sim Gau(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

The sampling design is a two phase design in which n_1 units are selected and observed at the first phase and the remaining $n_2 = n - n_1$ units are selected at the second phase. The sample sizes of each phase are fixed as n_1 = sample size of the first phase, n_2 = sample size of the second phase and $n = n_1 + n_1$, final sample size. The final data is $d = (d_1, d_2) = ((s_1, \mathbf{y}_{s_1}), (s_2, \mathbf{y}_{s_2}))$.

The predictor used for T is the best unbiased predictor (BUP), which with the specified model is $\hat{T} = \mathrm{E}(T|d)$.

2.1 Optimal Adaptive Strategy

The optimal adaptive design would select the second phase sampling units/locations s_2, under the BUP used and the model given above to minimize

$$\begin{aligned} g_{s_2}(s_1, \mathbf{y}_{s_1}) &= \mathrm{E}\{[\hat{T}(d_1, s_2, \mathbf{Y}_{s_2}) - T(\mathbf{Y})]^2 \mid d_1\} \\ &= \int (T - \hat{T})^2 f(\mathbf{y}_{\bar{s}_1} | d_1; \phi) d\mathbf{y}_{\bar{s}_1} \end{aligned} \tag{1}$$

(using the result of Zacks 1969 with the notation of Thompson and Seber 1996). The optimal design in n phases can be produced by applying this result recursively, but the computations are correspondingly more complex.

Because $\hat{T} = \mathrm{E}(T \mid d)$, therefore

$$g_{s_2}(s_1, \mathbf{y}_{s_1}) = \mathrm{E}[\mathrm{Var}(T|d)|d_1]$$

The optimal s_2, denoted by s_2^*, needs to satisfy the following condition

$$g_{s_2^*}(s_1, \mathbf{y}_{s_1}) = \min_{s_2 \subset \mathcal{S}_2} g_{s_2}(s_1, \mathbf{y}_{s_1})$$

where $\mathcal{S}_2$ is the set of all possible s_2. We should notice that $\mathcal{S}_2$ is a finite set since the population size is finite. However, the number of different possible s_2 is often very large.

For convenience we denote conditioning by subscripts with

$$\begin{aligned} \mathrm{E}_{d_1} &= \mathrm{E}(T|d_1) \\ \mathrm{Var}_{d_1}(T) &= \mathrm{E}\{[T - \mathrm{E}(T|d_1)]^2 | d_1\} = \mathrm{E}_{d_1}[T - \mathrm{E}_{d_1}(T)]^2 \end{aligned}$$

and

$$\mathrm{Var}(T|d) = \mathrm{Var}(T|d_1, d_2) = \mathrm{Var}_{d_1}(T|d_2).$$

Then

$$\begin{aligned} g_{s_2}(s_1, \mathbf{y}_{s_1}) &= \mathrm{E}_{d_1}[\mathrm{Var}_{d_1}(T|d_2)] \\ &= \mathrm{Var}_{d_1}(T) - \mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2)] \end{aligned}$$

Since the value of $\mathrm{Var}_{d_1}(T)$ does not depend on s_2, so the minimization of Equation (1) is equivalent to finding s_2^* which can maximize the value of $\mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2)]$ among all possible s_2. This means we shall find s_2^* such that

$$\mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2^*)] = \max_{s_2 \subset \mathcal{S}_2} \mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2)], \quad (2)$$

The calculation of $\mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2)]$ for the different possible combinations of s_2 can be done by the following results.

Let $\mathbf{y}_{s_1} = \{y_i\}_{i \in s_1}$, $\mathbf{Y}_{s_2} = \{Y_i\}_{i \in s_2}$, $\mathbf{Y}_{\bar{s}} = \{Y_i\}_{i \in \bar{s} = \{1,\ldots,N\} \setminus s}$ and $\mathbf{Y}_s = (\mathbf{y}_{s_1}, \mathbf{Y}_{s_2})$ represent the associated values of the units in the final sample. It is still a random vector since the real values of the second phase sample are unknown. Also, let $\mathbf{w}_{s_1}$, $\mathbf{W}_{s_2}$, $\mathbf{W}_s$ and $\mathbf{W}_{\bar{s}}$ represent the associated vectors of $\mathbf{y}_{s_1}$, $\mathbf{Y}_{s_2}$, $\mathbf{Y}_s$ and $\mathbf{Y}_{\bar{s}}$ transformed by the natural log function.

The conditional mean vector and variance-covariance matrix of $\mathbf{W}_{\bar{s}}$ given the first phase sampling data $d_1 = (s_1, \mathbf{w}_{s_1})$ and the potential second phase data $d_2 = (s_2, \mathbf{W}_{s_2})$ are $\boldsymbol{\nu} = \{\nu_i\}_{i \in \bar{s}}$ and $\boldsymbol{\Gamma} = \{\gamma_{ij}\}_{i,j \in \bar{s}}$, given by

$$\boldsymbol{\nu} = \boldsymbol{\mu}_{\bar{s}} + \boldsymbol{\Sigma}_{\bar{s}s}\boldsymbol{\Sigma}_{ss}^{-1}(\mathbf{W}_s - \boldsymbol{\mu}_s) \quad (3)$$

$$\boldsymbol{\Gamma} = \boldsymbol{\Sigma}_{\bar{s}\bar{s}} - \boldsymbol{\Sigma}_{\bar{s}s}\boldsymbol{\Sigma}_{ss}^{-1}\boldsymbol{\Sigma}_{s\bar{s}} \quad (4)$$

(cf. Arnold 1990, p. 214). Note $\boldsymbol{\Gamma}$ is fixed once s_2 is given, but $\boldsymbol{\nu}$ is still a random vector since $\mathbf{W}_{s_2}$ (and $\mathbf{Y}_{s_2}$) remains unknown.

Furthermore, let

$$\boldsymbol{\nu}^{\bar{s}_1} = \boldsymbol{\mu}_{\bar{s}_1} + \boldsymbol{\Sigma}_{\bar{s}_1 s_1}\boldsymbol{\Sigma}_{s_1 s_1}^{-1}(\mathbf{w}_{s_1} - \boldsymbol{\mu}_{s_1}) = \begin{pmatrix} \nu_{s_2}^{\bar{s}_1} \\ \\ \nu_{\bar{s}}^{\bar{s}_1} \end{pmatrix},$$

$$\boldsymbol{\Gamma}^{\bar{s}_1} = \boldsymbol{\Sigma}_{\bar{s}_1\bar{s}_1} - \boldsymbol{\Sigma}_{\bar{s}_1 s_1}\boldsymbol{\Sigma}_{s_1 s_1}^{-1}\boldsymbol{\Sigma}_{s_1\bar{s}_1} = \begin{pmatrix} \boldsymbol{\Gamma}_{s_2 s_2}^{\bar{s}_1} & \boldsymbol{\Gamma}_{s_2\bar{s}}^{\bar{s}_1} \\ \\ \boldsymbol{\Gamma}_{\bar{s}s_2}^{\bar{s}_1} & \boldsymbol{\Gamma}_{\bar{s}\bar{s}}^{\bar{s}_1} \end{pmatrix}$$

and decompose $\boldsymbol{\Sigma}_{\bar{s}s}\boldsymbol{\Sigma}_{ss}^{-1} = (\mathbf{A}_{s_1}, \mathbf{A}_{s_2})$. The dimension of $\mathbf{A}_{s_1}$ and $\mathbf{A}_{s_2}$ are $(N-n)\times n_1$ and $(N-n)\times n_2$ respectively. Also let $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}_{s_2 s_1}\boldsymbol{\Sigma}_{s_1 s_1}^{-1}$

It can be shown, by the properties of multivariate lognormal random vector and its associated Gaussian random vector, that

$$\begin{aligned} &\mathrm{Var}_{d_1}[\mathrm{E}_{d_1}(T|d_2)] \\ &= \sum_{\bar{s}_1}\sum_{\bar{s}_1} \exp\left(\eta_i + \frac{\delta_{ii}}{2}\right)\exp\left(\eta_j + \frac{\delta_{jj}}{2}\right)(e^{\delta_{ij}} - 1) \end{aligned} \quad (5)$$

where

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\nu}_{s_2}^{\bar{s}_1} \\ \\ \boldsymbol{\mu}_{\bar{s}} + (\mathbf{A}_{s_1} + \mathbf{A}_{s_2}\boldsymbol{\Sigma}')(\mathbf{w}_{s_1} - \boldsymbol{\mu}_{s_1}) + \frac{\mathrm{diag}(\boldsymbol{\Gamma})}{2} \end{pmatrix} \quad (6)$$

$$\boldsymbol{\Delta} = \begin{pmatrix} \boldsymbol{\Gamma}_{s_2 s_2}^{\bar{s}_1} & (\mathbf{A}_{s_2}\boldsymbol{\Gamma}_{s_2 s_2}^{\bar{s}_1})^T \\ \\ \mathbf{A}_{s_2}\boldsymbol{\Gamma}_{s_2 s_2}^{\bar{s}_1} & \mathbf{A}_{s_2}\boldsymbol{\Gamma}_{s_2 s_2}^{\bar{s}_1}\mathbf{A}_{s_2}^T \end{pmatrix} \quad (7)$$

All the information needed to calculate $\boldsymbol{\eta}$ and $\boldsymbol{\Delta}$ can be obtained after we have

d_1 : the first phase data, $(s_1, \mathbf{y}_{s_1})$.
s_2 : the second phase sampling location.

The calculation of Equation (5) then can be done by applying (6) and (7) with any value of d_1, s_2. Also we can conclude that the minimization of $g_{s_2}(s_1, \mathbf{y}_{s_1})$ depends on the $\mathbf{y}$-value of the first phase sample via $\boldsymbol{\eta}$ since $w_i = \log(y_i)$, $\forall i \in s_1$. This means that if we want to find s_2 which can minimize equation (1), we need to use an adaptive design.

2.2 Optimal Conventional Strategy

The optimal conventional design, on the other hand, would select s_2 to minimize the mean square error without taking the first-phase observations $\mathbf{y}_{s_1}$ into account.

Let

$$g_{s_2}^*(s_1) = \mathrm{E}[(\hat{T} - T)^2 | s_1, s_2]$$

then

$$\begin{aligned} g_{s_2}^*(s_1) &= \mathrm{E}[(\hat{T} - T)^2 | s_1, s_2] \\ &= \mathrm{E}[(\mathrm{Var}(T|d)|s_1, s_2] \\ &= \mathrm{Var}(T|s_1, s_2) - \mathrm{Var}[\mathrm{E}(T|d)|s_1, s_2] \\ &= \mathrm{Var}(T) - \mathrm{Var}[\mathrm{E}(T|s, \mathbf{Y}_s)] \end{aligned} \quad (8)$$

The object for the optimal conventional two phase design is to find a set $s_2 \in \mathcal{S}_2$ to minimize equation Equation (8).

The value of Var(T) is irrelevant to the minimization of $g^*_{s_2}(s_1)$ since it is the same for every different s_2. The only term in equation (8) we need to calculate in order to obtain the optimal s_2 is Var[E($T|s, \mathbf{Y}_s$)]. Using an approach similar to that in 2.1, the calculation of Var[E($T|s, \mathbf{Y}_s$)] is

$$\begin{aligned} &\text{Var}[\text{E}(T|s, \mathbf{Y}_s)] \\ = &\sum_i \sum_j \exp(\eta_i + \frac{\delta_{ii}}{2}) \exp(\eta_j + \frac{\delta_{jj}}{2})(e^{\delta^{ij}} - 1) \end{aligned} \tag{9}$$

where δ_{ij} is the i_{th} row, j_{th} column entry of the variance-covariance Δ, given by

$$\Delta = \begin{pmatrix} \mathbf{\Sigma}_{ss} & \mathbf{\Sigma}_{s\bar{s}} \\ \mathbf{\Sigma}_{\bar{s}s} & \mathbf{\Sigma}_{\bar{s}s}\mathbf{\Sigma}^{-1}_{ss}\mathbf{\Sigma}_{s\bar{s}} \end{pmatrix} \tag{10}$$

and η_i is the i_{th} entry of $\boldsymbol{\eta}$, given by

$$\boldsymbol{\eta} = \begin{pmatrix} \boldsymbol{\mu}_s \\ \\ \boldsymbol{\mu}_s + \frac{\text{diag}(\mathbf{\Gamma})}{2} \end{pmatrix} \tag{11}$$

$\mathbf{\Gamma}$ is the conditional variance-covariance matrix of $\mathbf{W}_{\bar{s}}$ given $\mathbf{W}_s$ which can be obtained by (4). We can use (9), (10) and (11) to calculate the values of Var[E($T|s, \mathbf{Y}_s$)] for different s_2 in order to find the optimal s_2 which can minimize Equation (8) under conventional sampling strategy.

3. Optimal Selection of Spatial Sampling Units

In this section, the conventional and adaptive optimal selection of sites will be illustrated in a spatial setting. For the example a square spatial study region contains $N = 81$ potential sampling sites, located at the cross points of a rectangular grid. The object is to select a sample of $n = 18$ of the sites, and predict the total $T = \sum_{i=1}^{81} Y_i$. The realizations of the 81 population values $\mathbf{Y}$ were produced using the lognormal model with a Gaussian-shaped spatial covariance function (cf. Cressie 1993, p.85) given by

$$\text{C}(\mathbf{h}) = \sigma^2 e^{-\|\mathbf{h}\|^2/a^2}, \mathbf{h} \in \mathbf{R}^d.$$

Thus the model is

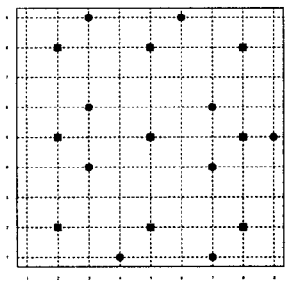
$$\log Y_i = W_i, \forall i = 1, ..., N$$

$$W_i \sim \text{N}(\mu, \sigma^2), \forall i = 1, ..., N$$

$$\text{Cov}(W_i, W_j) = \sigma^2 e^{-\|\mathbf{h}\|^2/a^2},$$

where $\| \mathbf{h} \|$ is the Euclidean distance between sites i and j. Parameter values $\mu = 0$, $\sigma^2 = 1$, and $a = 3$ were used.

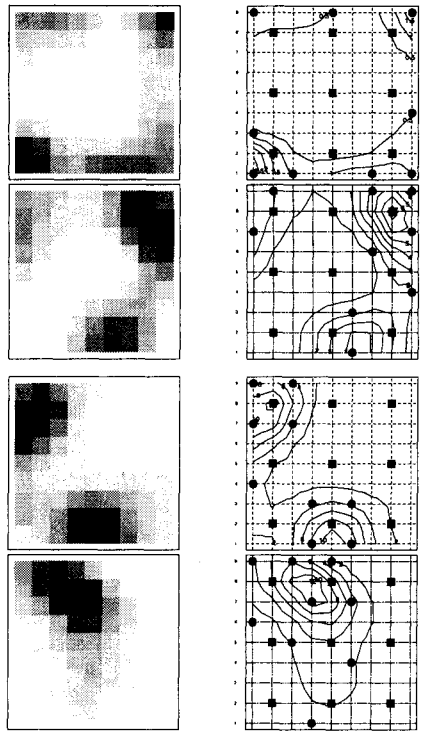
Figure 1: The Second Phase Sampling Locations under Optimal Conventional Design. (■ : First phase sample. • : Second phase sample.)



The first-phase sample is systematic and remains the same from realization to realization. The choice of second phase sampling locations is independent of $\mathbf{y}_{s_1}$ and takes into account only the locations s_1 already selected. Figure 1 shows the optimal conventional second phase sample selection. The result shows the tendency to spread the units out evenly. Only the discrete partitioning of the study region into units prevents a completely systematic arrangement.

The optimal adaptive second phase sampling locations, on the other hand, depend on the observed value of the first phase sample, $\mathbf{y}_{s_1}$. Thus the selection of second phase units varies from realization to realization. Figure 2 shows different optimal sampling locations selected with several different realizations. In Figure 2, the sample selections are superimposed on the realizations. Note that the observer does not see the entire realization, but only those values at the sample sites. The choice of the second-phase sample is based on observed values only at the first-phase sites. The superimposition of the sample selections on the underlying realizations illustrates the tendency of the optimal selection procedure to

Figure 2: The Second Phase Sampling Location (•) under Optimal Adaptive Design.



place sampling sites near high observed values and far from low observed values, while still striving to spread the sample out in the study region.

4. Relative Efficiency

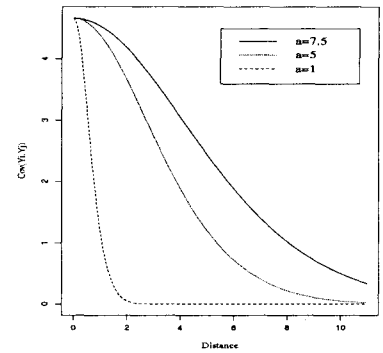
The spatial model of the proceeding section will be used in this section to compare the relative efficiencies of the optimal adaptive strategy (OA) with the optimal conventional strategy (OC). For the comparisons in this section, the square spatial study region contains $N = 49$ potential sampling locations on a rectangular grid. A sample of $n = 12$ sites are to be selected for the purpose of predicting the total $T = \sum_{i=1}^{49} Y_i$. For the two-phase strategies, the sample sizes are $n_1 = 4$ and $n_2 = 8$. The smaller population and sample sizes are used because of the large number of computations involved in the comparisons.

The relative efficiency of strategy 1 to strategy 2 is defined as the ratio of the mean square prediction error $\mathrm{E}(\hat{T} - T)^2$ obtained with strategy 2 to that obtained with strategy 1, so that a value greater than one indicates strategy 1 is the more efficient. Mean square prediction error was estimated with simulation by producing K realizations of the model and selection procedure and computing

$$\mathrm{E}(T - \hat{T})^2 = \frac{1}{K} \sum_{j=1}^{K} (T_j - \hat{T}_j)^2$$

The potential efficiency gain of the adaptive strategy over the conventional strategies is based on taking advantage of spatial dependencies in the population. The optimal conventional strategy uses the spatial structure only through the mean and covariance functions, while the adaptive strategy uses the spatial structure more fully. The shape of the covariance function in relation to the parameter a is illustrated in Figure 3.

Figure 3: The covariance $\mathrm{Cov}(Y_i, Y_j)$ with different values of a

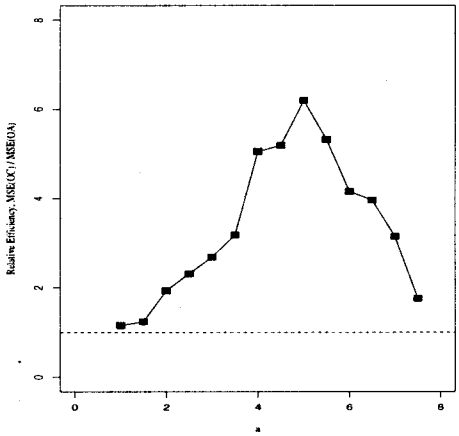


The parameter a is related to the range within which the spatial covariance function has the most influence. Specifically, a gives the spatial distance at which the covariance function is decreasing most rapidly. A small value of a corresponds to a covariance function approaching zero rapidly, so that observations at different sampling sites are approximately independent. With independence between units, one sampling design should be as good as another, so that relative efficiency should approach 1 as a decreases. As a increases, the covariance function becomes increasingly flat so that the correlation between the values in the study region approaches one. With the Y_i values perfectly correlated, the observed sites will provide a perfect prediction of the unobserved values. Thus, with any of the designs the mean square error will approach zero as a increases. Thus one would expect to see the maximum advantage of the adaptive strategy over the conventional ones for some intermediate value of a.

For values of a from 1 to 7.5, the computed relative efficiency of the optimal adaptive strategy to the optimal conventional strategy is shown in Figure 4. There are 4,000 different realizations generated for

each a and the experimental mean relative efficiency calculated out of the 4,000 different realizations is connected by the solid line. Of the computed values, the highest efficiency of the adaptive strategy relative to the others was obtained at the intermediate value $a \doteqdot 5$.

Figure 4: The relative efficiency of the optimal adaptive (OA) to the optimal conventional (OC) design with different values of the covariance parameter a.



5. Discussion

It is unlikely that a population model can be specified exactly in practice. However, the optimal sampling strategy under a Bayesian population model with a known prior distribution of the parameters can be developed in a similar manner as which in the given model situation. The reason is that the predictive inference of the pure prediction problem, with a known parameter, and the Bayesian prediction problem, with a known prior distribution, depend on the same conditional distribution given the data. For example, if we assume that the mean vector follows a multivariate prior distribution, that is with $\mathbf{Y} = (Y_1, Y_2, ..., Y_N)'$ and $\mathbf{W} = (W_1, W_2, ..., W_N)'$ with $W_i = \log Y_i$, $\forall i = 1, ..., N$, given $\mathbf{b}$,

$$\mathbf{W} \,|\, \mathbf{b} \;\sim\; Gau(\mathbf{b}, \mathbf{V}).$$

and the prior distribution of $\mathbf{b}$ is

$$\mathbf{b} \;\sim\; Gau(\boldsymbol{\mu}, \boldsymbol{\Gamma}).$$

It can be shown (most easily using the moment generating function) that under this model the unconditional distribution of $\mathbf{W}$ is

$$\mathbf{W} \;\sim\; Gau(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

where

$$\boldsymbol{\Sigma} = \boldsymbol{\Gamma} + \mathbf{V}$$

Since the unconditional distribution of $\mathbf{W}$ is normal, the unconditional distribution of $\mathbf{Y}$ is again lognormal, and the derivations of all the conditional distributions already obtained apply.

The procedure for a more complicated and practical Bayesian population model, under which the marginal distribution of the population variable of interest is not of closed form, is under investigation by the authors.

6. References

Arnold, S.F. (1990). *Mathematical Statistics.* Englewood Cliffs, New Jersey: Prentice-Hall.

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā A* **31**, 441–454.

Cressie, N. (1993). *Statistics for Spatial Data*, 2nd edition. New York: Wiley.

Guttorp, P., Le, N.D., Sampson, P.D., and Zidek, J.V. (1993). Using entropy in the redesign of an environmental network. *In* G.P. Patil and C.R. Rao (Eds), *Multivariate Environmental Statistics*, pp. 175–202. New York: North Holland/Elsevier Science Publishers.

Matérn, B. (1986). *Spatial Variation*, 2nd edition. Berlin: Springer-Verlag.

Ripley, B.D. (1976). The second-order analysis of stationary point process. *Journal of Applied Probability* **13**, 255-266.

Sacks, J., and Schiller, S. (1988). Spatial designs. *In* S.S. Gupta and J.O. Berger (Eds), *Statistical Decision Theory and Related Topics IV, Vo. 2*, pp. 385–389. New York: Springer.

Thompson, S.K. (1988). Adaptive sampling. *Proceedings of the Section on Survey Research Methods of the American Statistical Association*, 784–786.

Thompson, S.K. (1990). Adaptive cluster sampling. *Journal of the American Statistical Association* **85**, 1050–1059.

Thompson, S.K., and Seber, G.A.F. (1996). *Adaptive Sampling.* New York: Wiley.

Zacks, S. (1969). Bayes sequential designs of fixed size samples from finite populations. *Journal of the American Statistical Association* **64**, 1342–1349.