Journal of Statistical Research of Iran
Vol. **2** No. 1 pp. 41-42

مجله‌ی پژوهش‌های آماری ایران
دوره‌ی ۲، شماره‌ی ۱، بهار و تابستان ۱۳۸۴، صص ۶۵-۶۶

Determination of Optimal Sampling Design for Spatial Data Analysis

M. Jafari Khaledi* and F. Rivaz

Extended Abstract. Inferences for spatial data are affected substantially by the spatial configuration of the network of sites where measurements are taken. Consider the following standard data-model framework for spatial data. Suppose that a continuous, spatially-varying quantity, Z, is to be observed at a predetermined number, n, of points $S_n = \{t_1, \ldots, t_n\}$ in a region of interest. Let $Z = (Z(t_1), \ldots, Z(t_n))$ represent the observations taken at these points. These observations are modeled statistically as a spatially incomplete sample of one realization of a random field $\{Z(t) : t \in D\}$. Assume further that the random field's mean is of the form $E\{Z(t)\} = \sum_{i=1}^{p} \beta_i f_i(t)$, where f_i's are known function of observed covariates. We assume that the covariance function parameter θ in $C(s, u; \theta) = \mathrm{Cov}(Z(s), Z(u))$ is known.

Under the model just described, the generalized least square estimation and its variance are given by

$$\hat{\beta}_{gls} = (X'\Sigma_\theta^{-1}X)^{-1}X'\Sigma_\theta^{-1}Z, \quad \mathrm{Var}(\hat{\beta}_{gls}) = (X'\Sigma_t heta^{-1}X)^{-1},$$

where X is a $n \times p$ matrix with ijth element $f_{j-1}(t_i)$ and assumed to have full column rank; Σ_θ is $n \times n$ matrix with ijth element $C(t_i, t_j; \theta)$. The best (in the sense of minimizing the prediction error variance) linear unbiased predictor (BLUP) of $Z(t_0)$, the realized but unobserved value of Z at an arbitrary point, and prediction variance is given by

$$\begin{aligned}
\hat{Z}(t_0) &= f'(t_0)\hat{\beta}_{gls} + C_\theta'\Sigma_\theta^{-1}(Z - X\hat{\beta}_{gls}), \\
\sigma^2(t_0, S_n) &= E\{Z(t_0) - \hat{Z}(t_0)\}^2, \\
&= V^2(t_0, S_n) + \phi'(t_0, S_n)(X'\Sigma_\theta^{-1}X)\phi(t_0, S_n),
\end{aligned}$$

Two types of design questions can be asked under this model:

* Corresponding author.

1. What is the optimal sampling design for estimating regression parameters?

2. What is the optimal sampling design for spatial prediction?

By defining the criterion function as

$$\Phi\{\mathrm{Var}(\hat{\beta}_{gls})\} = |X'\Sigma_\theta^{-1}X|^{-1}$$

we show that the optimal sampling design for estimating regression parameter is

$$S_n^* = \arg\min_{s_n \in D} |X'\Sigma_\theta^{-1}X|^{-1}.$$

Also, with respect to the following mean and maximum prediction variance criteria

$$V(S_n) = \frac{1}{|D|}\int_D \sigma^2(x, S_n)\, dx \quad H(S_n) = \max_{x\in D} \sigma^2(x, S_n),$$

the optimal sampling design for spatial prediction are respectively given by

$$S_n^* = \arg\min_{S_n \in D} V(S_n), \quad S_n^* \arg\min_{S_n \in D} H(S_n).$$

In general, it is impossible to find the optimal sampling design of continuous design region D. Suppose now that D is approximated by a fine grid with size N, denoted by D_N. Thus, the construction of an optimal spatial sampling design of size n reduces to finding the best n sites from among all $\binom{N}{n}$ possible sampling plans. When $\binom{N}{n}$ is large, the naive optimization can be computationally prohibitive. In this case, we proposed appropriate algorithms to search an approximately optimal design among all possible design on the fine grid.

Keywords. spatial data; spatial sampling design; optimality.

Magid Jafari Khaledi
Department of Statistics,
Faculty of Basic Science,
Tarbiat Modares University,
Tehran, Iran.
e-mail: *jafari_m@modares.ac.ir*

Firoozeh Rivaz
Department of Statistics,
Faculty of Basic Science,
Tarbiat Modares University,
Tehran, Iran.
e-mail: *rivaz@modares.ac.ir*

The full version of the paper, in Persian, appears on pages 47-54.