The Korean Communications
in Statistics Vol. 12 No. 2, 2005
pp. 539-544

One-step Least Squares Fitting of Variogram

Hyemi Choi[1)]

Abstract

In this paper, we propose the one-step least squares method based on the squared differences to estimate the parameters of the variogram used for spatial data modelling, and discuss its asymptotic efficiency. The proposed method does not require to specify *lags of interest* and partition lags, so that we can delete the subjectiveness and ambiguity originated from the lag selection in estimating spatial dependence.

Keywords : Asymptotic efficiency, Generalized least squares estimation, Variogram

1. Introduction

Spatial statistics has recently received an increasing attention. Applications of spatial statistical models are extremely numerous and diverse. These include image analysis, atmospheric sciences and geostatistics, among others. Spatial observations from nearby locations are often related. Hence, the estimation of covariance parameters is an important problem for spatial processes. The covariance structure of the spatial process is typically described in terms of its variogram.

A popular approach for estimating variogram parameters is the method of least squares model fitting, which is initially proposed in the geostatistical literature and then further modified and studied by Cressie (1985), Zhang *et al.* (1995), Genton (1998) and others. This method guarantees the conditionally negative definiteness property (Cressie (1993), Section 2.3.1) of the resulting variogram estimator and has a visual appeal that is similar to fitting a regression function to a scatter plot, which makes it very popular among practitioners.

The classical least squares methods consists of two steps such that fits a parametric variogram by minimizing the distance between a generic nonparametric variogram estimator and the parametric model at a *finite* number of particular lags, i.e. $h_1, \cdots, h_K$ by using various least squares methods. When data are irregularly spaced, which case frequently

1) Postdoctoral researcher, Division of Applied Mathematics, KAIST, Daejeon, 305-761, Korea
E-mail : hyemichoi@yahoo.com

happens in practice, the so-called K *tolerance* regions are considered. Each region consists of the location pairs whose separations are approximately equal to the lag of interest. The least squares principle is applied to the nonparametric estimator obtained in each of K regions. Journel and Huijbrets (1978) recommend that the number of distinct pairs in each region should be at least 30. Often the regions are chosen to be disjoint.

For the nonparametric estimator, the method suggested by Matheron and its modifications are commonly used, which are the (weighted) *average* of squared differences of a process after suitable mean adjustments. In this paper, we suppose that the mean of the observed spatial process is constant. It is clear from the description of the method that the properties of the fitted variogram depend on the particular distance function used to fit the parametric model. Recently, Lahiri *et al.*(2002) and Lee and Lahiri (2002) have shown that the generalized least squares (GLS) estimators and its variants of variogram parameters are asymptotically efficient among all least squares procedures.

In this paper, we propose the one-step least squares method based on the squared differences and discuss the one-step least squars estimators attaining the same asymptotic efficiency as GLS estimators. It may not be only computationally much simpler than the GLS method but does not need to specify *lags of interest* and partition lags, so that we can delete the subjectiveness and ambiguity in estimating spatial dependence, which is originated from the lag selection.

In Section 2, we describe the LSEs and some relevant results on variogram model fitting. The issue of large sample efficiencies of the LSEs is considered in Section 3. Section 4 discusses the optimal selection rule of lags of interest for the estimation of variogram parameters.

2. One-step Least Squares Method for Variogram Estimation

Let $\{Z(s) : s \in \mathbb{R}^d, d \geq 1\}$ be an intrinsically stationary random field, i.e. for all $s, h \in \mathbb{R}^d$ $E(Z(s) - Z(s+h)) = 0$ and $var(Z(s) - Z(s+h)) = var(Z(0) - Z(h))$. The function $2\gamma(h) \equiv var(Z(0) - Z(h))$ is called the variogram of the process $Z(\cdot)$. We assume that the true variogram of $Z(\cdot)$ lies in a parametric family $\{2\gamma(\cdot;\theta) : \theta \in \Theta \subset \mathbb{R}^q\}$ of valid variograms. Let $Z(s_1), \cdots, Z(s_n)$ denote the observations on the process $Z(\cdot)$ at sampling locations $s_1, \cdots, s_n$ in the sampling region $R \subset \mathbb{R}^d$. To reduce computational burden, we only consider n_λ terms of squared differences, denoted by Y_k of the observation pairs whose observed locations are not farther separated than a given constant $\lambda > 0$; that is, the squared differences $(Z(s_i) - Z(s_j))^2$ of the case $\|s_i - s_j\| \leq \lambda$ for a norm $\|\cdot\|$. Also, let l_k, $k = 1, ..., n_\lambda$ denote the lags $s_i - s_j$ corresponding to the squared difference $Y_k = (Z(s_i) - Z(s_j))^2$. Note that $n_\lambda = O(n)$ and $E(Y_i) = 2\gamma_i(\theta)$. Also let $Y = (Y_1, \cdots, Y_{n_\lambda})^t$

and $\gamma(\theta) = (\gamma_1(\theta), \cdots, \gamma_{n_\lambda}(\theta))^t$.

Now, we define the one-step LSEs of the variogram parameter θ minimizing certain quadratic distance measures between Y and $2\gamma(\theta)$. Let $M(\theta)$ be a $K \times K$ positive definite matrix for $q \leq K < n_\lambda$ and $P(\theta)$ an $n_\lambda \times K$ matrix of full rank. Throughout this paper, we drop the argument θ of the matrices such as M and P to simplify the notation. Then, a one-step LSE of θ corresponding to M and P is defined as

$$\hat{\theta}_{n,M,P} = argmin\{Q_n(\theta; M, P) : \theta \in \Theta\},$$

where $Q_n(\theta; M, P) = (Y - 2\gamma(\theta))^t P M^{-1} P^t (Y - 2\gamma(\theta))$. This method applies the LS principle directly to the mean-adjusted squared differences of process instead of the differences between the nonparametric estimator and parametric model at the given lags of interest. We may refer to the matrix P in the quadratic form as a projection of the n_λ -dimensional vector $Y - 2\gamma(\theta)$ onto K-dimensional space.

It may be seen that the one-step LSEs for properly chosen P is the same as classical LSEs. In fact, the classical LSEs using the Matheron's estimator as a generic nonparametric variogram estimator are a version of the one-step LSEs as follows: consider the Matheron's variogram estimator $2\hat{\gamma}$ with *lags of interest* $h_k, k = 1, \cdots, K$ taken to be the averages of the *squared differences* $(Z(s_i) - Z(s_j))^2$'s corresponding to the location pairs, i.e.,

$$2\hat{\gamma}(h_k) = \frac{1}{|N(h_k)|} \sum_{N(h_k)} (Z(s_i) - Z(s_j))^2,$$

where $N(h_k)$ denotes the set of the location pairs (s_i, s_j) belonging to the some specified disjoint tolerance region $T(h_k)$ in $\mathbb{R}^d$ around $h_k, k = 1, \cdots, K$. The Matheron's estimator $2\hat{\gamma}$ can be expressed as $P_0^t Y$, where $P_0 = A(A^t A)^{-1}$ for the $n_\lambda \times K$ matrix A having, as its (i,j)th entry, $a_{i,j} = 1(l_i \in T(h_j))$, where $1(\cdot)$ is the indicator function. Then, the *classical* LSEs based on Matheron's method corresponds to the minimizer of $Q_n(\theta; M, P_0)$ with $\gamma^*(\theta) \equiv (\gamma(h_1;\theta), \cdots, \gamma(h_K;\theta))^t$ in place of $P_0^t \gamma(\theta)$. The classical LS methods implicitly assumes that $P_0^t \gamma - \gamma^*(\theta)$ is negligible. For the matrix $M = P_0 W P_0^t$, the one with $W = I_K$, the identity matrix of order K, is the OLS estimator θ based on Matheron's method. Choosing W as the covariance matrix of $2\hat{\gamma}$, we obtain the GLS estimator of θ. Similarly, choosing W to be a suitable diagonal matrix, we can obtain the various WLS estimators proposed by Cressie (1985), Zhang *et al.* (1995) and Genton (1997). Especially, the

approximated weighted least squares (AWLS) estimator of Cressie (1985) corresponds to the one with $W = diag(4\gamma^2(h_1;\theta),\cdots,4\gamma^2(h_K;\theta))$.

3. Asymptotic Properties of One-step Least Squares Estimators

In this section, we will describe the asymptotic normality of the one-step least squares estimator. Large sample properties of the classical LSEs have been recently obtained by Lahiri *et al.* (2002). The asymptotic properties of the one-step LSE are given in the following without mathematical details, which can be derived from the related results Theorems 3.1-3.3 in Lahiri *et al.* (2002). For this, let $\Gamma(\theta)$ denote the $n_\lambda \times q$ matrix with (i,j)th element $-(2\partial\gamma_i/\partial\theta_j)$ and let $\Sigma(\theta) = cov(Y)$.

If for sufficiently large n and all $\theta_1 \neq \theta_2$, $(2\gamma(\theta_1) - 2\gamma(\theta_2))^t PP^t(2\gamma(\theta_1) - 2\gamma(\theta_2)) > 0$, then, under the compactness of Θ and the smoothness conditions on P and M, the one-step LSE $\hat{\theta}_{M,P}$ is consistent.

Suppose that the matrices M and P satisfy that $P^t(Y - 2\gamma(\theta)) \rightarrow N(0, \Sigma_P)$ for some nonsingular matrix $\Sigma_P = \lim_{n\to\infty} P^t\Sigma P$ and for sufficiently large n, $\frac{1}{\sqrt{n}}\Gamma(\theta)^t PM^{-1}P^t\Gamma(\theta)$ is positive definite for all θ. The application of Theorems 3.2 and 3.3 of Lahiri *et al.* (2002) yields that $\sqrt{n}(\hat{\theta}_{n,M,P} - \theta) \rightarrow N(0, \Sigma_{M,P}(\theta))$, where $\Sigma_{M,P}(\theta) = \lim_{n\to\infty} n\hat{\Sigma}_{M,P}(\theta)$ with $\hat{\Sigma}^{-1}_{M,P}(\theta) = \Gamma^t PM^{-1}P\Gamma(\Gamma^t PM^{-1}P^t\Sigma PM^{-1}P^t\Gamma)^{-1}\Gamma^t PM^{-1}P^t\Gamma$.

4. Efficiency Considerations

In this section we address the relative efficiencies of least squares variogram parameter estimators. We consider the LSEs with M and P satisfying the conditions given in Section 3 only. We call a one-step LSE $\hat{\theta}_{n,M_0,P}$ *asymptotically efficient* if, for any weighting matrix M, the difference $\Sigma_{M,P}(\theta) - \Sigma_{M_0,P}(\theta)$ of the asymptotic covariance matrices of the LSEs $\hat{\theta}_{n,M,P}$ and $\hat{\theta}_{n,M_0,P}$ is non-negative definite for all $\theta \in \Theta$.

Let M_G and P_G denote P^tGP and $G^{-1}\Gamma$ for $n_\lambda \times n_\lambda$ matrix G respectively. Now, the covariance matrix $\hat{\Sigma}_{M_\Sigma,P}$ of $\hat{\theta}_{n,M_\Sigma,P}$ is expressed as $(\Gamma^t P(P^t\Sigma P)^{-1}P^t\Gamma)^{-1}$ and hence, by Theorem 4.1 of Lahiri *et al.* (2002) we have $\hat{\Sigma}_{M,P} - \hat{\Sigma}_{M_\Sigma,P}$ is nonnegative definite, denoted by $\hat{\Sigma}_{M,P} - \hat{\Sigma}_{M_\Sigma,P}[\geq]0$, for any nonsingular matrix M, which implies that the one-step LSE $\hat{\theta}_{n,M_\Sigma,P}$ corresponding to the GLS estimator for given P is asymptotically efficient among all

one-step LSEs with P. However, although it is optimal from the statistical point of view, a computation of the GLS estimator can be difficult in practice.

Next, we note that for any nonsingular G

$$\begin{aligned}\hat{\Sigma}_{M_G,P_G} &= (\Gamma^t G^{-1}\Gamma)^{-1}\Gamma^t G^{-1}\Sigma G^{-1}\Gamma(\Gamma^t G^{-1}\Gamma)^{-1} \\ &= (\Gamma^t P_G(P_G^t\Sigma P_G)^{-1}P_G^t\Gamma)^{-1},\end{aligned}$$

which equals $\hat{\Sigma}_{M_\Sigma,P_\Sigma}$. Thus it follows that $\hat{\Sigma}_{M,P_G}-\hat{\Sigma}_{M_G,P_G}[\geq]0$ for any nonsingular $K\times K$ matrix M and hence $\hat{\theta}_{n,M_G,P_G}$ for each G is efficient among $\hat{\theta}_{n,M,P_G}$'s.

Now, we compare the *efficiencies* of $\hat{\theta}_{n,M_G,P}$'s with given G based on the inverse covariance matrix. The asymptotic inverse covariance matrix $\hat{\Sigma}^{-1}_{M_G,P}(\theta)$ of $\hat{\theta}_{n,M_G,P}$ can be represented as $\Gamma^t\Pi_\Sigma[\Pi_G[P]\Gamma]\Gamma$, where $\Pi_A[B]=B(B^tAB)^{-1}B^t$. By using the fact that $X^t\Pi_A[B]X$ is maximized when the column space $Col(X)$ generated by X is a subspace of $Col(AB)$, we see that $\hat{\Sigma}^{-1}_{M_G,P}(\theta)$ is maximized when $\Gamma\in Col(\Sigma\Pi_G[P]\Gamma)$ and hence $\Gamma\in Col(\Pi_G[P]\Gamma)$, which is attained in the case that $P=G^{-1}\Gamma$, that is $P=P_G$. Thus we have that $\hat{\Sigma}_{M_G,P}-\hat{\Sigma}_{M_G,P_G}[\geq]0$ and hence $\hat{\theta}_{n,M_G,P_G}$ is *efficient* among $\theta_{n,M_G,P}$'s. In other words, for $M=M_G$, the P_G is an optimal *projection* matrix of the n_λ-dimensional vector $Y-2\gamma(\theta)$ onto K-dimensional space, $q\leq K<n_\lambda$. Recall that the classical LSEs using nonparametric estimator based on squared differences Y can be considered as the one-step LSE and note that P in such classical LSE depends on *lags of interest* that are subjectively selected. Then, we may say by the optimality of $\hat{\theta}_{n,M_G,P_G}$ that P_G provides the optimal lag selection rule.

In addition, the Gauss-Markov theorem gives the nonnegativity of $\hat{\Sigma}_{M_G,P_G}-\hat{\Sigma}_{M_\Sigma,P_\Sigma}$. It implies that $\hat{\theta}_{n,M_\Sigma,P_\Sigma}$ is optimal among $\hat{\theta}_{n,M_G,P_G}$ for any nonsingular G. If there exists arbitrary $q\times q$ matrix R such that $\Gamma^tG^{-1}\Sigma G^{-1}\Gamma=\Gamma^tG^{-1}\Gamma R$, then $\hat{\Sigma}_{M_G,P_G}=\hat{\Sigma}_{M_\Sigma,P_\Sigma}$. the $\hat{\theta}_{n,M_G,P_G}$ for such G attains the same asymptotic efficiency as the GLS estimator and may be computationally much simpler.

References

[1] Cressie, N. (1985). Fitting variogram models by weighted least squares, *Journal of the International Association for Mathematical Geology*, Vol. 17, 563-586.

[2] Cressie, N. (1993). *Statistics for Spatial Data, revised edition*, Wiley.

[3] Cressie, N. and Hawkins, D. M. (1980). Robust estimation of the variogram, I, *Journal of the International Association for Mathematical Geology*, Vol. 12, 115-125.

[4] Genton, M.G. (1998). Highly robust variogram estimation, *Mathematical Geology*, Vol. 30, 213-221.

[5] Genton, M.G. (1998). Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure, *Mathematical Geology*, Vol. 30, 323-345.

[6] Lahiri, S.N, Lee, Y.D. and Cressie, N. (2002). On asymptotic distribution and asymptotic efficiency of least squares estimators of spatial variogram parameters, *Journal of Statistical Planning and Inference,* Vol. 103, 65-85.

[7] Lee, Y.D. and Lahiri, S.N. (2002). Least Squares Variogram fitting by Spatial Subsampling, *The Journal of the Royal Statistical Society*, Vol. B64, 837-854.

[8] Zhang, X., van Eijkeren, J. and Heemink, A. (1995). On the weighted least-squares method for fitting a semivariogram model, *Computers & Geosciences*, Vol. 21, 605-608.

[Received March 2005, Accepted July 2005]