Biometrika (2005), **92,** 4, *pp.* 921–936

Printed in Great Britain

Towards reconciling two asymptotic frameworks in spatial statistics

By HAO ZHANG

Department of Statistics, Washington State University, Pullman, Washington 99164-3144, U.S.A.

zhanghao@wsu.edu

AND DALE L. ZIMMERMAN

Department of Statistics and Actuarial Science, University of Iowa, Iowa City, Iowa 52242, U.S.A.

dzimmer@stat.uiowa.edu

Summary

Two asymptotic frameworks, increasing domain asymptotics and infill asymptotics, have been advanced for obtaining limiting distributions of maximum likelihood estimators of covariance parameters in Gaussian spatial models with or without a nugget effect. These limiting distributions are known to be different in some cases. It is therefore of interest to know, for a given finite sample, which framework is more appropriate. We consider the possibility of making this choice on the basis of how well the limiting distributions obtained under each framework approximate their finite-sample counterparts. We investigate the quality of these approximations both theoretically and empirically, showing that, for certain consistently estimable parameters of exponential covariograms, approximations corresponding to the two frameworks perform about equally well. For those parameters that cannot be estimated consistently, however, the infill asymptotic approximation is preferable.

Some key words: Asymptotic normality; Consistency; Increasing domain asymptotics; Infill asymptotics; Maximum likelihood estimation; Spatial covariance.

1. Introduction

Spatially referenced data are usually positively spatially correlated, observations from nearby sites tending to be more alike than observations from distant sites. It is standard practice to model the data's variance and correlation structure through a parametric covariance function, or covariogram, and then estimate these parameters, by the method of maximum likelihood, say. For purposes of making inferences about the covariogram's parameters, knowledge of the asymptotic properties of parameter estimators is useful, mainly because one hopes that the asymptotic results will yield useful approximations to finite-sample properties. However, the applicability of asymptotics to spatial data is complicated by the fact that there are two quite different asymptotic frameworks to which one can appeal: increasing domain asymptotics, in which the minimum distance between sampling points is bounded away from zero and thus the spatial domain of observation is unbounded, and infill asymptotics, in which observations are taken ever more densely in a fixed and bounded domain.

Not surprisingly, the asymptotic behaviour of spatial covariance parameter estimators can be quite different under the two frameworks. It is known, for example, that some covariance parameters are not consistently estimable under infill asymptotics (Ying, 1991; Stein, 1999, p. 110; Zhang, 2004), whereas these same parameters are consistently estimable and their maximum likelihood estimators are asymptotically normal, subject to some regularity conditions, under increasing domain asymptotics (Mardia & Marshall, 1984). Furthermore, there are cases in which a parameter is consistently estimable under both asymptotic frameworks, but the convergence rates are different (Chen et al., 2000).

Typically in practice, spatial data are observed at a finite number of points with no intention or possibility of taking more observations, and it is not clear which asymptotic framework to appeal to. Stein (1999) gives a cogent argument for using infill asymptotics if interpolation of the spatial process is the ultimate goal. An alternative approach is to choose a framework on the basis of how well the asymptotic distributions of estimators of parameters of interest approximate the finite-sample distributions of those estimators. The purpose of this paper is to investigate and compare the quality of these approximations.

2. Review of some asymptotic results

Consider a spatial process $X(s)$ that is observed on a set of n points D_n, with $D_1 \subset D_2 \subset \ldots \subset R^d$, and whose distribution depends on the parameter $\theta \in R^p$, where p and d are fixed positive integers. Let $L_n(\theta)$ be the likelihood function of θ given the observations $\{X(s): s \in D_n\}$. Then a maximum likelihood estimator of θ is any value $\hat{\theta}_n$ that maximises $L_n(\theta)$. If $X(s)$ is Gaussian, Mardia & Marshall (1984) showed that, under an increasing domain asymptotic framework and subject to some regularity conditions, $\hat{\theta}_n$ is approximately normally distributed with mean θ and covariance matrix $I_n^{-1}(\theta)$, where

$$I_n(\theta) = -E\{\partial^2 \log L_n(\theta)/\partial\theta\,\partial\theta'\}.$$

One of the regularity conditions is that the diagonal elements of $I_n^{-1}(\theta)$ converge to 0 as $n \to \infty$.

The available results under infill asymptotics are considerably narrower in scope than for increasing domain asymptotics. Consider a stationary, zero-mean, Gaussian process that has an exponential covariogram, that is

$$E\{X(s)\} = 0, \quad \text{cov}\{X(s), X(s+h)\} = \theta_1 \exp(-\theta_2 h) \quad (s \in R, h \geqslant 0). \tag{2·1}$$

When this process is observed in the unit interval, Ying (1991) showed that, as $n \to \infty$,

$$\surd n(\hat{\theta}_1\hat{\theta}_2 - \theta_1\theta_2) \to N\{0, 2(\theta_1\theta_2)^2\}, \tag{2·2}$$

in distribution, where $\hat{\theta}_i$ is the maximum likelihood estimator of θ_i $(i = 1, 2)$. Furthermore, if θ_2 is fixed at any value $\tilde{\theta}_2$, then the estimator $\hat{\theta}_1 = \arg\max L_n(\theta_1, \tilde{\theta}_2)$ satisfies

$$\surd n(\hat{\theta}_1 - \theta_1) \to N\{0, 2(\theta_1\theta_2/\tilde{\theta}_2)^2\}, \tag{2·3}$$

in distribution. In particular, if θ_2 is known and $\tilde{\theta}_2 = \theta_2$, the limiting variance in (2·3) is $2\theta_1^2$. However, the individual parameters θ_1 and θ_2 are not consistently estimable under the infill asymptotic framework; see Zhang (2004) for more general results.

When there are measurement errors, the infill asymptotic behaviour of the maximum likelihood estimators is somewhat different. Chen et al. (2000) showed that, for a zero-mean Gaussian process on the unit interval having an exponential covariogram with a nugget effect, that is

$$\operatorname{cov}\{X(s), X(s+h)\} = \begin{cases} \theta_0 + \theta_1, & \text{if } h = 0, \\ \theta_1 \exp(-\theta_2 h), & \text{if } h > 0, \end{cases} \tag{2·4}$$

the maximum likelihood estimators $\hat{\theta}_i$ $(i = 0, 1, 2)$ satisfy

$$\begin{pmatrix} n^{1/2}(\hat{\theta}_0 - \theta_0) \\ n^{1/4}(\hat{\theta}_1\hat{\theta}_2 - \theta_1\theta_2) \end{pmatrix} \to N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\theta_0^2 & 0 \\ 0 & 4(2\theta_0)^{1/2}(\theta_1\theta_2)^{3/2} \end{pmatrix}\right\}, \tag{2·5}$$

in distribution. Also, if the inverse range parameter θ_2 is known, then the maximum likelihood estimators of θ_0 and θ_1 satisfy

$$\begin{pmatrix} n^{1/2}(\hat{\theta}_0 - \theta_0) \\ n^{1/4}(\hat{\theta}_1 - \theta_1) \end{pmatrix} \to N\left\{\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\theta_0^2 & 0 \\ 0 & 4(2\theta_0)^{1/2}\theta_1^{3/2}\theta_2^{-1/2} \end{pmatrix}\right\}, \tag{2·6}$$

in distribution. Note that, in this case, $\hat{\theta}_1\hat{\theta}_2$ and $\hat{\theta}_1$ converge at the rate of $n^{-1/4}$ instead of $n^{-1/2}$ as in the previous case.

Analogous results to Ying (1991) and Chen et al. (2000) are not available for two-dimensional, and higher, isotropic cases, though Ying (1993) established the infill asymptotic distribution of maximum likelihood estimators of the parameters of a separable exponential covariogram in higher dimensions. Zhang (2004) showed that not all parameters in a Matérn covariance function are consistently estimable under infill asymptotics, and he identified one parametric function that is consistently estimable when the dimension $d = 1$, 2 or 3.

The increasing domain asymptotic distribution of $(\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2)$ for the case of model (2·4) does not appear to be available in the literature, but we will derive it in § 3·2.

We now review a general result for martingales, which is useful for establishing properties of maximum likelihood estimators under any asymptotic framework. Let $\mathscr{F}_n$ be the σ-algebra generated by $\{X(s): s \in D_n\}$. For any $i = 1, \ldots, p$, it can be shown that $\{\partial \log L_n(\theta)/\partial\theta_i, \mathscr{F}_n, n \geqslant 1\}$ is a martingale. Thus, under a wide range of circumstances, we have

$$\mathscr{I}_n^{-1/2}(\theta)\partial \log L_n(\theta)/\partial\theta \to N(0, I),$$

in distribution, where $\mathscr{I}_n(\theta)$ is the conditional information matrix, as given by (6.2) of Hall & Heyde (1980), and I is the identity matrix; see for example Proposition 6.1 of Hall & Heyde (1980). The asymptotic normality of the maximum likelihood estimator of θ can be established using the asymptotic results above and the first-order Taylor expansion of $\partial \log L_n(\xi)/\partial\xi$ about θ; see Crowder (1976) and equation (6.4) of Hall & Heyde (1980). This approach underlies how the asymptotic distributions of maximum likelihood estimators are established in the previously cited works.

There is a simple case in which the limiting distributions are the same under both frameworks. Let $\{X(s): s \in R^d\}$, be a stationary, zero-mean, Gaussian process with covariogram $C(h) = (1/\theta)\rho(h)$, where $\rho(h)$ is a known correlogram. If we observe $X_n = \{X(s_i), i = 1, \ldots, n\}$, it is easily verified that the Fisher information and conditional information coincide and equal $I_n(\theta)=n/(2\theta^2)$. Furthermore, $\partial \log L_n(\theta)/\partial\theta$ satisfies Assumptions 1 and 2 of Hall & Heyde (1980). Proposition 6.1 of Hall & Heyde (1980) then implies that

$$I_n^{-1/2}(\theta)\partial \log L_n(\theta)/\partial\theta = I_n^{-1/2}(\theta)\left(\frac{n}{2\theta}-\tfrac{1}{2}X_n'\Gamma_n^{-1}X_n\right)\to N(0,1),$$

in distribution, or equivalently

$$\sqrt{n}\left(\frac{1}{n}X_n'\Gamma_n^{-1}X_n-\frac{1}{\theta}\right)\to N(0,2/\theta^2),$$

in distribution, where $(\Gamma_n)_{ij}=\rho(s_i-s_j)$. This result holds as long as $n\to\infty$, regardless of framework. The key here is that the Fisher information matrix does not depend on X_n, and therefore behaves in the same way under both frameworks. However, when the correlogram has a parameter to be estimated, the information matrix may behave differently under the two frameworks. For example, the diagonal elements of the inverse matrix of the Fisher information matrix may not go to 0 as $n\to\infty$ under infill asymptotics. This difference may be the driving force behind the different results under the two frameworks.

We finish this section with a general discussion about the behaviour of maximum likelihood estimators under infill asymptotics. Let the process $\{X(s): s\in D\}$ be Gaussian with a mean and covariogram that depend on a vector parameter ϕ, where D is a bounded infinite subset of R^d for some dimension d. Following Stein (1999, p. 163), we say that a function $h(\phi)$ is microergodic if, for all ϕ and $\tilde{\phi}$ in the parameter space, $h(\phi)\neq h(\tilde{\phi})$ implies that the two measures $P(\phi)$ and $P(\tilde{\phi})$ are orthogonal, where $P(\phi)$ denotes the Gaussian measure corresponding to the parameter ϕ. Microergodicity is necessary but not sufficient for the existence of a consistent estimator. Let us partition $\phi=(\phi_1,\phi_2)$ such that ϕ_1 has only microergodic elements and ϕ_2 has only non-microergodic elements. In addition, we assume that, for any $\phi=(\phi_1,\phi_2)$ and $\tilde{\phi}=(\tilde{\phi}_1,\tilde{\phi}_2)$, $P(\phi)$ and $P(\tilde{\phi})$ are equivalent if and only if $\phi_1=\tilde{\phi}_1$. We will now argue that, when the observations become dense in D, the maximum likelihood estimators of ϕ_i $(i=1,2)$ have the following properties under regularity conditions: $\hat{\phi}_1$ is asymptotically normal; $\hat{\phi}_2$ converges in probability or almost surely to the maximum likelihood estimator of ϕ_2 when ϕ_1 is known and the process is observed everywhere in D. This limit is nondegenerate.

We will now indicate why these results are plausible. Let $P(\phi_1,\phi_2)$ be the Gaussian measure on the σ-algebra generated by $\{X(s): s\in D\}$, and let $\phi_{i,0}$ denote the true value of ϕ_i $(i=1,2)$. Since the two measures $P(\phi_1,\phi_2)$ and $P(\phi_{1,0},\phi_{2,0})$ are equivalent if $\phi_1=\phi_{1,0}$ and orthogonal otherwise, we have that, with $P(\phi_{1,0},\phi_{2,0})$-probability 1, the Radon–Nikodym derivative

$$\rho(\phi_1,\phi_2)=\frac{dP(\phi_1,\phi_2)}{dP(\phi_{1,0},\phi_{2,0})}$$

is equal to 0 if $\phi_1\neq\phi_{1,0}$, and is strictly positive for any ϕ_2 if $\phi_1=\phi_{1,0}$. If the process is observed everywhere in D, the Radon–Nikodym derivative is a likelihood, and consequently the maximum likelihood estimator of ϕ_1 is the degenerate variable $\phi_{1,0}$, and the maximum likelihood estimator of ϕ_2 maximises $dP(\phi_{1,0},\phi_2)/dP(\phi_{1,0},\phi_{2,0})$, which is the maximum likelihood estimator of ϕ_2 when $\phi_{1,0}$ is known. Denote this estimator of ϕ_2 by $\hat{\phi}_{2,\infty}$.

Given observations $X(s)$ for s in a finite subset D_n of D, let $P_n(\phi_1, \phi_2)$ denote the restriction of $P(\phi_1, \phi_2)$ to the σ-algebra generated by $\{X(s): s \in D_n\}$. The maximum likelihood estimators $\hat{\phi}_{1,n}$ and $\hat{\phi}_{2,n}$ maximise the density

$$\rho_n(\phi_1, \phi_2) = \frac{dP_n(\phi_1, \phi_2)}{dP_n(\phi_{1,0}, \phi_{2,0})}.$$

If the subsets D_n increase to D, that is $D_n \subset D_{n+1}$ and $\cup_{n=1}^{\infty} D_n = D$, then $\rho_n(\phi_1, \phi_2)$ converges almost surely to $\rho(\phi_1, \phi_2)$; see for example Theorem 1 of Gihman & Skorohod (1974, p. 442). Consequently, $\hat{\phi}_{1,n}$ and $\hat{\phi}_{2,n}$ converge in probability or almost surely, depending on the assumed regularity conditions, to $\phi_{1,0}$ and $\hat{\phi}_{2,\infty}$, respectively.

The asymptotic normality of $\hat{\phi}_{1,n}$ can be established under conditions similar to those given by Crowder (1976).

An explicit expression for $\hat{\phi}_{2,\infty}$ does not always exist; however, it can be given in some special cases. For example, consider the Ornstein–Uhlenbeck process $\{X(t): 0 \leqslant t \leqslant T\}$ that has mean 0 and satisfies the stochastic differential equation

$$dX(t) = -\phi_2 X(t)dt + \surd(2\phi_1)dB(t) \quad (\phi_1 > 0, \phi_2 > 0),$$

where $B(t)$ is a Brownian motion. This process has an exponential covariogram $(\phi_1/\phi_2)\exp(-\phi_2 h)$, for $h > 0$ (Karatzas & Shreve, 1991, p. 358). For any finite $T > 0$, any fixed $\phi_1 > 0$ and any $\phi_2 > 0$, the probability measure restricted to $\sigma\{X(t), 0 \leqslant t \leqslant T\}$, denoted by P^X, is equivalent to the probability measure restricted to $\sigma\{B(t), 0 \leqslant t \leqslant T\}$, denoted by P^B, and the density is

$$dP^B/dP^X = \exp\left[(2\phi_1)^{-1}\left\{\phi_2 \int_0^T X(t)dX(t) + \tfrac{1}{2}\phi_2^2 \int_0^T X(t)^2 dt\right\}\right]$$

(Liptser & Shiryayev, 1977, Theorem 7.7, p. 248). Then the likelihood function is $1/(dP^B/dP^X)$, which is maximised by

$$\hat{\phi}_{2,\infty} = -\frac{\int_0^T X(t)dX(t)}{\int_0^T X(t)^2 dt}. \tag{2·7}$$

Therefore, if the process is observed continuously on $[0, T]$, the maximum likelihood estimator of ϕ_2 is given explicitly by (2·7).

3. Asymptotic results for two specific models

3·1. *Model* 1

Consider a Gaussian process having mean 0 and nuggetless exponential covariogram (2·1). We assume initially that the process is observed at points $s_i = \delta i$, for $i = 0, \ldots, n$ and some fixed constant $\delta > 0$ not depending on n. The asymptotic distribution of the maximum likelihood estimators of θ_i $(i = 1, 2)$ can be established easily under this increasing domain framework. Note that the time series $Y_i = X(s_i)$ has a power correlation function $R(k) = \rho^{|k|}$, for $\rho = \exp(-\theta_2\delta)$, and therefore follows a Gaussian AR(1) model:

$$Y_i - \rho Y_{i-1} = \varepsilon_i \quad (i = 1, 2, \ldots, n),$$

where the $\{\varepsilon_i\}$ are independently and identically distributed as $N(0, \eta)$ and $\eta = \theta_1(1 - \rho^2)$.

It is well known that the maximum likelihood estimator of $(\eta, \rho)'$ is asymptotically normal, that is

$$\sqrt{n}\left\{\begin{pmatrix}\hat{\eta}\\ \hat{\rho}\end{pmatrix} - \begin{pmatrix}\eta\\ \rho\end{pmatrix}\right\} \to N(0, W),$$

in distribution, where $W = \mathrm{diag}(2\eta^2, 1-\rho^2)$.

The maximum likelihood estimator of $\theta = (\theta_1, \theta_2)'$ thus has the following asymptotic distribution:

$$\sqrt{n}(\hat{\theta}_n - \theta) \to N(0, AWA'), \tag{3·1}$$

in distribution, where

$$A = \begin{pmatrix}\partial\theta_1/\partial\eta & \partial\theta_1/\partial\rho\\ \partial\theta_2/\partial\eta & \partial\theta_2/\partial\rho\end{pmatrix}.$$

After some calculation, we obtain

$$AWA' = \begin{pmatrix}2\theta_1^2(1+\rho^2)/(1-\rho^2) & -2\theta_1/\delta\\ -2\theta_1/\delta & (1-\rho^2)/(\delta\rho)^2\end{pmatrix}. \tag{3·2}$$

Furthermore,

$$\sqrt{n}(\hat{\theta}_1\hat{\theta}_2 - \theta_1\theta_2) \to N(0, \sigma^2), \tag{3·3}$$

in distribution, for

$$\sigma^2 = 2(\theta_1\theta_2)^2 + \{(1-\rho^2)(\delta\rho)^{-1} - 2\theta_2\rho\}^2\theta_1^2(1-\rho^2)^{-1}. \tag{3·4}$$

We now calculate explicitly the Fisher information matrix given a finite sample. The (i, j)th element of the Fisher information matrix is one half of the trace of $V^{-1}V_iV^{-1}V_j$, where $V_i = \partial V/\partial\theta_i$ $(i, j = 1, 2)$ and V is the covariance matrix of $Y_0, \ldots, Y_n$ (Mardia & Marshall, 1984). We see that, for this particular model, $V_1 = (1/\theta_1)V$, and after some calculation we find that the diagonal elements of $V^{-1}V_2$ are $2\delta\rho^2/\{\theta_1(1-\rho^2)\}$, except for the first and last ones, which are $\delta\rho^2/\{\theta_1(1-\rho^2)\}$. The diagonal elements of $(V^{-1}V_2)^2$ are $2\delta^2\rho^2(1+\rho^2)/(1-\rho^2)^2$, except for the first and last ones, which are $\delta^2\rho^2(1+\rho^2)/(1-\rho^2)^2$. The information matrix is therefore

$$I_n(\theta) = \frac{1}{2\theta_1^2}\begin{pmatrix}n+1 & \theta_1 h\\ \theta_1 h & \theta_1^2 g\end{pmatrix},$$

where $h = 2\delta n\rho^2/(1-\rho^2)$ and $g = 2\delta^2 n\rho^2(1+\rho^2)/(1-\rho^2)^2$. The inverse is

$$I_n^{-1}(\theta) = c^{-1}\begin{pmatrix}\theta_1^2 g & -\theta_1 h\\ -\theta_1 h & n+1\end{pmatrix}, \tag{3·5}$$

where

$$c = \{(n+1)g - h^2\}/2 = \frac{n\delta^2\rho^2}{1-\rho^2}\left(n + \frac{1+\rho^2}{1-\rho^2}\right).$$

Note that c is dominated by the first term of the right-hand side. It follows that $nI_n^{-1}(\theta)$ converges to the covariance matrix AWA' in (3·1) as $n \to \infty$, and thus we can use I_n^{-1} to approximate the covariance matrix of the $\hat{\theta}_i$'s. This is expected in the light of the general result of Mardia & Marshall (1984).

Now suppose instead that the sampling points are given by $s_i = i/n$ $(i = 0, \ldots, n)$. The inverse information matrix in this case is still given by (3·5), but its asymptotic behaviour is quite different since $\delta = 1/n$ depends on n here, which affects h and g. It can be shown that

$$h = \frac{2\rho^2}{1-\rho^2} = \frac{n}{\theta_2} + o(n), \quad g = \frac{2\rho^2(1+\rho^2)}{n(1-\rho^2)^2} = \frac{n}{\theta_2^2} + o(n), \quad c = \frac{1+\theta_2}{2\theta_2^2} n + o(n).$$

It follows that

$$\lim_{n\to\infty} I_n^{-1}(\theta) = \frac{\theta_2}{1+\theta_2} B,$$

where

$$B = \begin{pmatrix} 2\theta_1^2/\theta_2 & -2\theta_1 \\ -2\theta_1 & 2\theta_2 \end{pmatrix}.$$

We see that the diagonal elements of the inverse information matrix do not converge to 0 as $n \to \infty$. In addition, the limit of $I_n(\theta)$ is singular. Thus, some of the basic assumptions of Mardia & Marshall (1984) do not hold and consequently, under infill asymptotics, there is no theoretical basis for using the normal distribution to approximate the distribution of $\hat{\theta}_i$ $(i = 1, 2)$.

Although approximations to the covariance matrix of $\hat{\theta}$ based on either (3·1) or the inverse information matrix are inappropriate when $\delta = 1/n$, it is nonetheless of some interest to see how the two approximations compare to each other. It can be shown that, when $\delta = 1/n$ in (3·2),

$$\lim_{n\to\infty} AWA'/n = B.$$

Thus, if we use $I_n^{-1}(\theta)$ or AWA'/n to approximate the variance matrix of $(\hat{\theta}_1, \hat{\theta}_2)$, the difference is approximately a multiplicative factor $\theta_2/(1+\theta_2)$ for a large sample. This factor is closer to 1 when θ_2 is large, which corresponds to weaker correlation. Therefore, while neither of these two approximations is appropriate under infill asymptotics, the difference between them is not substantial unless the correlation is very strong.

Finally let us consider a different parameterisation by letting $\phi_1 = \theta_1\theta_2$ and $\phi_2 = \theta_2$. Then, under both asymptotic frameworks, ϕ_1 is consistently estimable and its maximum likelihood estimator is asymptotically normal. If we calculate the Fisher information matrix for the new parameters, the variance of $\hat{\phi}_1$ as given by the inverse information matrix is $2\phi_1^2/n + o(1/n)$ when $\delta = 1/n$ (Abt & Welch, 1998), which agrees with the infill asymptotic result (2·2). Given observations at finitely-many uniformly spaced points in $[0, 1]$, we therefore have three distinct approximations to the distribution of $\hat{\phi}_1$, namely that given by (3·3) and (3·4), which was derived under an increasing domain framework, that given by (2·2), and the finite sample Fisher information matrix; the latter two were derived under infill asymptotics. We have just noted that the difference between the second and third approximations is negligible when the sample size n is large. Furthermore, it is easily shown that (3·4) converges to $2\phi_1^2$ when $\delta = 1/n$. Thus, the increasing domain approximation based on (3·4), which would seem to be inappropriate under infill sampling, will nevertheless be asymptotically equivalent to the other two approximations. Therefore, given a large sample from this particular model, the three approximations to the finite sample distribution of $\hat{\phi}_1$ would be virtually identical.

The infill asymptotic distribution of $\hat{\phi}_2$ coincides with that of (2·7), which will be shown in § 4 to be right-skewed and which therefore differs from the asymptotic normal distribution of $\hat{\phi}_2$ under the increasing domain framework.

3·2. *Model* 2

As our second model we consider a zero-mean Gaussian process having the exponential covariogram with a nugget effect. This process can be written as a sum of two independent Gaussian processes,

$$Y(s) = X(s) + W(s),$$

where $X(s)$ is the Gaussian process of Model 1 and $W(s)$ is Gaussian white noise, independent of $X(s)$, which accounts for measurement error. The covariogram of $Y(s)$ is given by (2·4), where the nugget effect θ_0 is the variance of $W(s)$ and θ_1 is the variance of $X(s)$. Recall that θ_1 and θ_2 are not consistently estimable under infill asymptotics, but that $(\theta_0, \theta_1\theta_2)'$ is consistently estimable and its maximum likelihood estimator is asymptotically normal; see (2·5).

First we establish the limiting distributions of the maximum likelihood estimators under increasing domain asymptotics. When $Y(s)$ is observed at δi for some fixed $\delta > 0$ $(i = 0, 1, \ldots)$, the resulting time series $Y_i = Y(\delta i)$ has been studied previously; see for example Gingras & Masry (1988) and Pagano (1974). However, this literature apparently does not include explicit asymptotic distributions of the maximum likelihood estimators.

Suppose that we observe $Y_0, \ldots, Y_n$. Note that Y_i has the spectral density

$$f(\lambda) = \frac{1}{2\pi}\left(v_0 + \frac{v_1}{1 + \rho^2 - 2\rho\cos\lambda}\right), \tag{3·6}$$

where $v_0 = \theta_0$, $\rho = \exp(-\delta\theta_2)$ and $v_1 = (1 - \rho^2)\theta_1$. Write $v_2 = \theta_2$. It follows from a well-known result (Rosenblatt, 1985, Theorems 3, 4) that, for the maximum likelihood estimator $\hat{v}$ of $v = (v_0, v_1, v_2)'$,

$$\sqrt{n}(\hat{v} - v) \to N(0, E^{-1}), \tag{3·7}$$

in distribution, where E is the matrix with $(i + 1, j + 1)$th $(i, j = 0, 1, 2)$ element

$$E_{ij} = \frac{1}{4\pi}\int_0^{2\pi} \frac{\partial \log f(\lambda)}{\partial v_i}\frac{\partial \log f(\lambda)}{\partial v_j}\, d\lambda. \tag{3·8}$$

Explicit expressions for the E_{ij}'s are obtained in the Appendix.

The maximum likelihood estimators for other parameterisations are also asymptotically normal. For example, for $\phi = (\theta_0, \theta_1\theta_2) = (v_0, v_1v_2/(1 - \rho^2))'$, we have

$$\sqrt{n}(\hat{\phi} - \phi) \to N(0, JE^{-1}J'), \tag{3·9}$$

in distribution, where $J = \partial\phi/\partial v$ is the 2×3 Jacobian matrix.

Although the limiting distributions for $\hat{\phi}$ under the two asymptotic frameworks, as given by (3·9) and (2·5), are both multivariate normal, the limiting covariance matrices seem quite different. Given a large but finite sample, the covariance matrix of $\hat{\phi}$ is approximately

$JE^{-1}J'/n$ according to (3·9), and diag$\{2\theta_0^2/n, 4(2\theta_0)^{1/2}(\theta_1\theta_2)^{3/2}/\sqrt{n}\}$ according to (2·5). We will show, however, that, when the sampling sites are $s_i = i/n$ $(i = 0, \ldots, n)$, these two covariance matrices agree asymptotically. To be more specific, let

$$A_n = JE^{-1}J'/n, \quad B_n = \text{diag}\{2\theta_0^2/n, 4(2\theta_0)^{1/2}(\theta_1\theta_2)^{3/2}/\sqrt{n}\}.$$

Then

$$B_n^{-1/2} A_n B_n^{-1/2} \to \text{diag}(1, 1),$$

as $n \to \infty$. Hence we can use either one to approximate the covariance matrix of $\hat{\phi}$. A proof of (3·10) is given in the Appendix.

4. A simulation study

Given a sequence of sets of sampling locations, the appropriate asymptotic sampling framework is obvious. However, in virtually every practical application there is only one sample and one set of sampling locations. Hence, it is often not clear which asymptotic framework and consequently which asymptotic results to employ. In order to identify typical finite-sample situations where each asymptotic approximation works or fails, we carried out a simulation study, which we now present.

The first process we simulated was Model 1, the stationary Gaussian process having mean 0 and an exponential covariogram (2·1), with θ_1 fixed at 1 and θ_2 equal to 4, 8 or 16. We took sampling sites to be equally spaced over [0, 1], and considered three sample sizes, namely 41, 81 and 161, although the results for the sample size 81 are not shown. Thus, 9 combinations of parameter value and sample size were considered. For each of these combinations, we simulated 1000 independent realisations of the Gaussian process at the sampling sites and then obtained maximum likelihood estimates using the Newton–Raphson algorithm described by Zhang (2004). In the numerical algorithm, we employed the parameterisation $\phi_1 = \theta_1\theta_2$ and $\phi_2 = \theta_2$. Thus ϕ_1 is microergodic and ϕ_2 is not.

There are three approximations to the finite-sample distribution of the maximum likelihood estimators of ϕ_1 and ϕ_2, as discussed in § 3. We compare the three approximations by comparing the quantiles of the approximate distributions with the empirical quantiles computed from the 1000 estimates. Figure 1 plots the $0{\cdot}05 + 0{\cdot}1(i-1)$, for $i = 1, \ldots, 10$, quantiles for parameter ϕ_1, where the horizontal axis is for the empirical quantiles and the vertical axis is for the quantiles of the three normal distributions corresponding to the finite-sample Fisher information, dotted line, the increasing domain approximation, circles, and the infill approximation, plus signs. We refrain from plotting additional quantiles in order to make the display more readable. We observe from Fig. 1 that all three approximations improve as the sample size n increases. Each of them fits the finite-sample distribution quite well when $n = 80$, not shown, or 160, while for $n = 40$ the finite-sample distribution of $\hat{\phi}_1$ is slightly to moderately right-skewed. For the microergodic parameter ϕ_1, there is little difference among the three approximations in all cases.

Figure 2 is a similar display for maximum likelihood estimates of ϕ_2, where the infill limit distribution is given by the distribution of $\hat{\phi}_{2,\infty}$ in (2·7). We used simulation to approximate this limit distribution. We simulated the Ornstein–Uhlenbeck process at $m = 5000$ points i/m $(i = 1, \ldots, m)$, which results in a first-order autoregressive time series

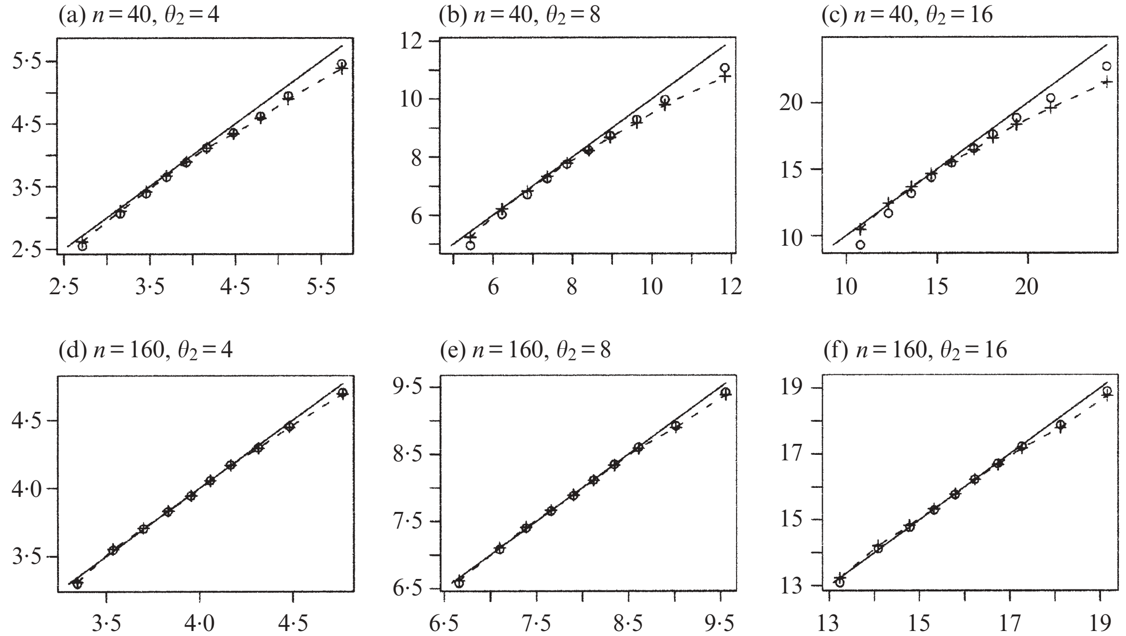

Fig. 1. Plots of quantiles of limit distributions according to the infill asymptotics, shown by plus signs, the increasing domain asymptotics, circles, and the finite-sample Fisher information, dotted line, against the empirical quantiles of estimates of $\phi_1 = \theta_1\theta_2$, where θ_1 is fixed at 1 and $\theta_2 = 4$, 8 and 16. Empirical quantiles were based on 1000 samples of size $n + 1$ from Model 1 at sites i/n $(i = 0, \ldots, n)$, for (a)–(c) $n = 40$ and (d)–(f) $n = 160$.

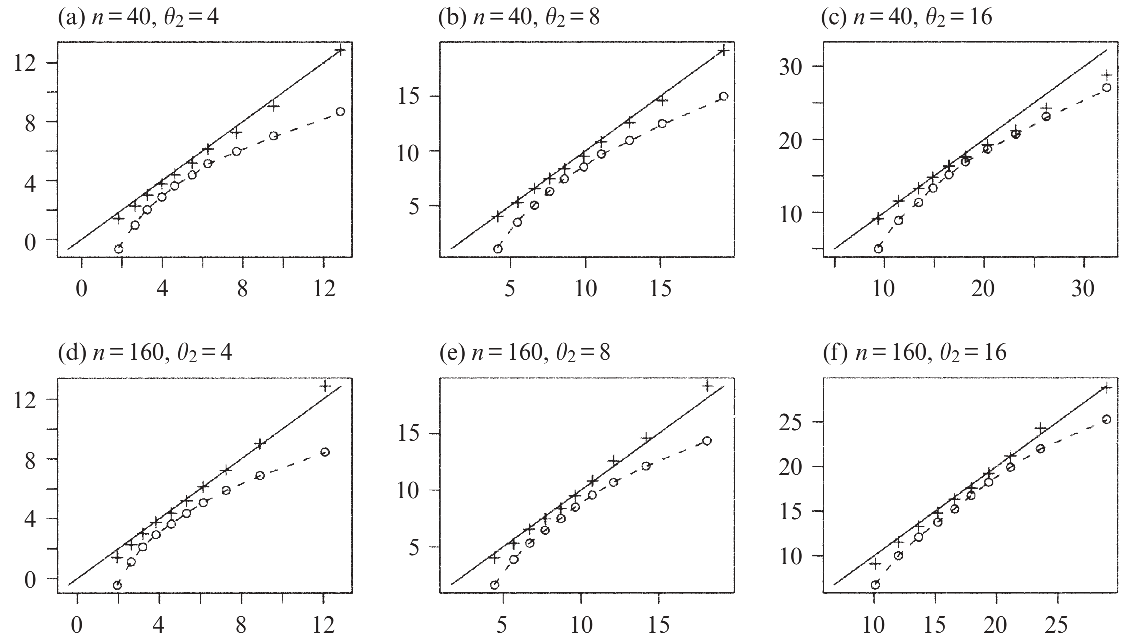

Fig. 2. Plots of quantiles of limit distributions according to the infill asymptotics, shown by plus signs, the increasing domain asymptotics, circles, and the finite-sample Fisher information, dotted line, against the empirical quantiles of estimates of θ_2, where θ_1 is fixed at 1 and $\theta_2 = 4$, 8 and 16. Empirical quantiles were based on 1000 samples of size $n + 1$ from Model 1 at sites i/n $(i = 0, \ldots, n)$, for (a)–(c) $n = 40$ and (d)–(f) $n = 160$.

$X(i/m)$. The quantity on the right-hand side of (2·7) is approximated by

$$-\frac{\sum_{i=1}^{m-1} X(i/m)[X\{(i+1)/m\} - X(i/m)]}{\sum_{i=1}^{m} X(i/m)^2/m}.$$

The simulation was repeated 1000 times to obtain an approximation to the distribution of $\hat{\phi}_{2,\infty}$. Figure 2 reveals that the finite-sample distribution of ϕ_2 is approximated very well by the infill limit distribution (2·7), even when $n=40$, but it is approximated very poorly by the normal approximations corresponding to the finite-sample Fisher information and increasing domain framework.

The design of our simulation study facilitates an investigation of not only infill asymptotic behaviour but also increasing domain asymptotic behaviour. For example, consider $n=40$ and $\phi_2=4$, and suppose we sample at $\{i/40 : i=0, \ldots, 40\}$. Adding 40 more sampling sites at $\{i/40 : i=41, \ldots, 80\}$ would correspond to increasing domain sampling, and the random variables generated at all 81 sites would be zero-mean Gaussian with $\mathrm{cov}(Y_i, Y_j)=\exp(-\phi_2|i-j|/40)$. Clearly, however, random variables generated at the 81 sites $\{i/80 : i=0, \ldots, 80\}$ from a zero-mean Gaussian process with variance 1 and inverse range parameter $2\phi_2$ have the same joint distribution. Therefore, we can mimic the increasing domain framework by fixing the domain and increasing the inverse range parameter. The increasing domain asymptotic behaviour of, say, $\hat{\phi}_2$, can be studied simply by diagonally examining plots in Fig. 2, from which we see that, when $\phi_2=4$, a sample size of 161 is not large enough for the increasing domain asymptotic distribution to approximate satisfactorily the finite-sample distribution of $\hat{\phi}_2$. In particular, the distribution of $\hat{\phi}_2$ is still skewed. In this case, the process is observed on the interval $[0, 4]$ and the effective range, that is the distance between two points at which the correlation is about 5%, is about 0·75. Therefore, although the spatial correlation in this case is rather weak, apparently it must be even weaker for the increasing domain asymptotic distributions to approximate the finite-sample distributions well.

Next, we simulated data from Model 2 at the same sites and with the same sample sizes. We fixed $\theta_1=2$ and took $\theta_0=1$, 2 and $\theta_2=4$, 8, 16. Thus, for this model we have 18 combinations of parameter value and sample size. For each combination, we simulated 1000 independent realisations of the Gaussian process. For each simulated realisation, we used a Fisher scoring algorithm (Mardia & Marshall, 1984) to obtain maximum likelihood estimates of θ_i $(i=0, 1, 2)$. As before, we partition the parameters into microergodic ones and a non-microergodic one by defining

$$\phi_0=\theta_0, \quad \phi_1=\theta_1\theta_2, \quad \phi_2=\theta_2.$$

Quantiles of fitted density and quantiles of estimates of θ_0 are plotted against each other in Fig. 3. We first note that θ_0 tends to be underestimated, and that the bias decreases when the sample size increases. When $n=40$, θ_0 is seriously underestimated, especially when the spatial correlation is weak. For example, when $n=40$, $\theta_0=1$ and $\theta_2=16$, the empirical results reveal a positive probability that the maximum likelihood estimator of θ_0 is exactly 0, which explains why some circles and plus signs appear vertically stacked for $n=40$; see the first row of Fig. 3. The three approximations become more similar when θ_2 decreases or when n increases. Depending on how strong the spatial correlation is, the sample size required for the three approximations to be close to each other may be quite large. For example, a sample size of 161 is sufficient for the three approximations to the

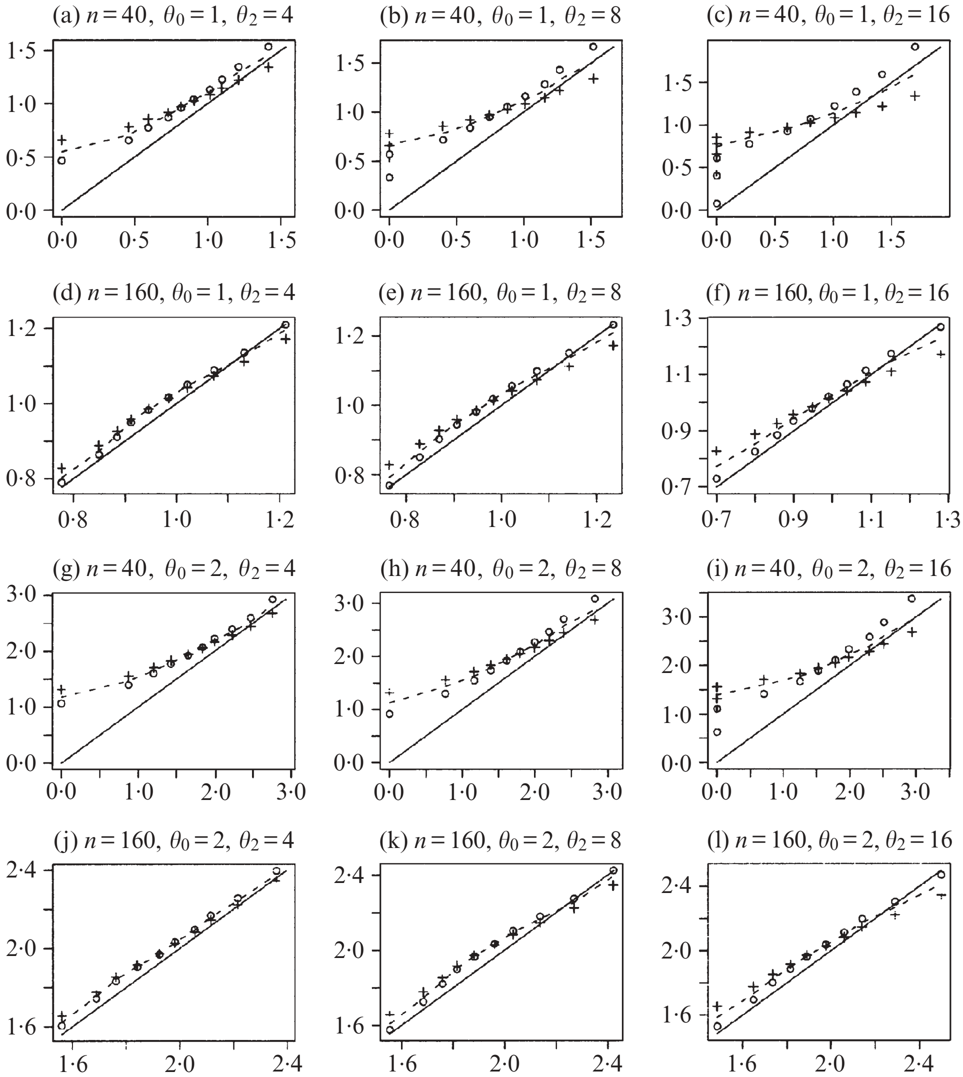

Fig. 3. Plots of quantiles of limit distributions according to the infill asymptotics, shown by plus signs, the increasing domain asymptotics, circles, and the finite-sample Fisher information, dotted line, against the empirical quantiles of the estimates of nugget effect θ_0, where θ_1 is fixed at 2 and $\theta_2 = 4$, 8 and 16. Empirical quantiles were based on 1000 samples of size $n+1$ from Model 2 at sites i/n $(i = 0, \ldots, n)$, for (a)–(c) $n = 40$ and $\theta_0 = 1$, (d)–(f) $n = 160$ and $\theta_0 = 1$, (g)–(i) $n = 40$ and $\theta_0 = 2$, (j)–(l) $n = 160$ and $\theta_0 = 2$.

distribution of $\hat{\theta}_0$ to be similar when $\theta_2 = 4$, but not when $\theta_2 = 16$. We also observe that the Fisher information appears to be a compromise between the infill asymptotic variance and the increasing domain asymptotic variance.

Figure 4 plots quantiles of the fitted density versus quantiles of estimates of ϕ_1. We see that, in all cases, the three approximations are very similar. However, none of them approximates the finite-sample distributions well, even when $n = 160$. This is because the distribution of $\hat{\phi}_1$ is right-skewed even when $n = 160$. Thus, the sample size required for any of the approximations to be satisfactory has to be much larger than 161. This is not surprising in the light of the slower convergence rate of $\hat{\phi}_1$ when a nugget effect is present, that is $n^{-1/4}$ rather than $n^{-1/2}$.

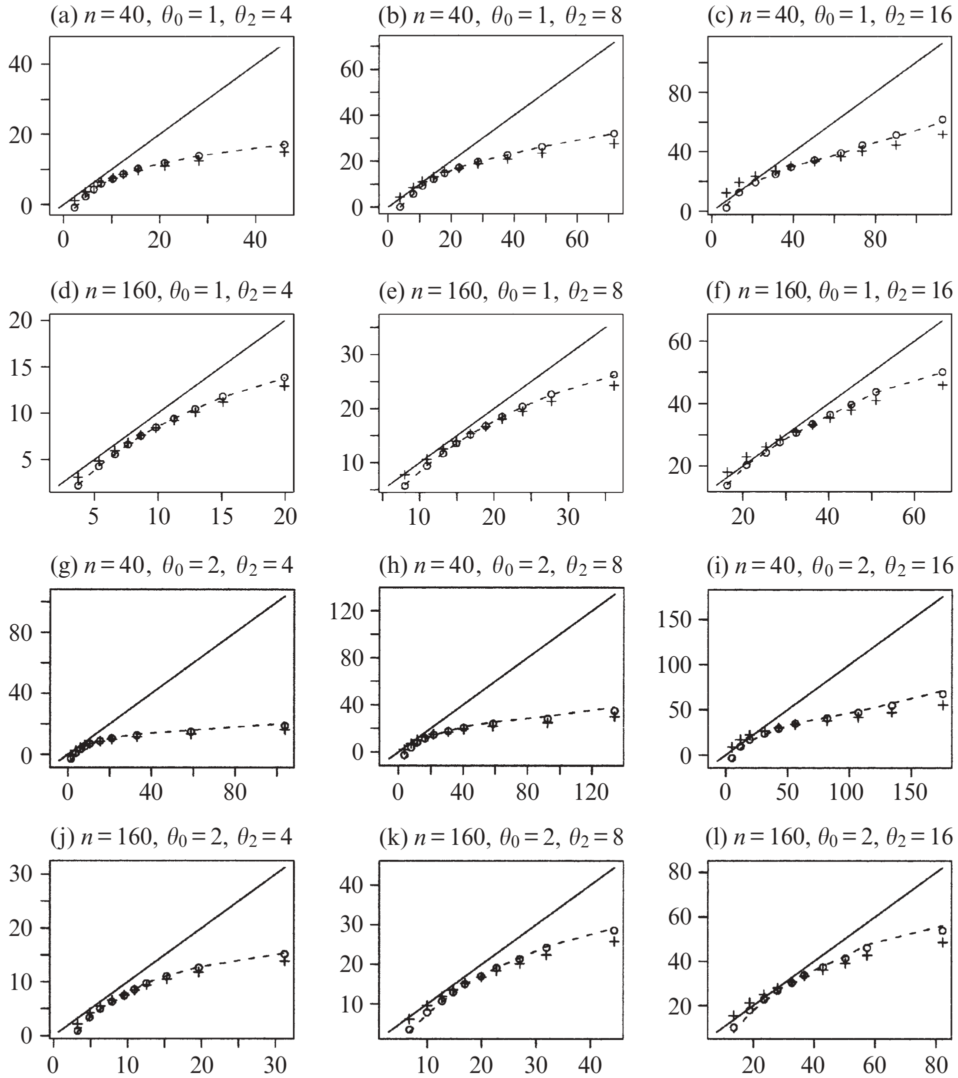

Fig. 4. Plots of quantiles of limit distributions according to the infill asymptotics, shown by plus signs, the increasing domain asymptotics, circles, and the finite-sample Fisher information, dotted line, against the empirical quantiles of the estimates of $\phi_1=\theta_1\theta_2$, where θ_1 is fixed at 2 and $\theta_2=4$, 8 and 16. Empirical quantiles were based on 1000 samples of size $n+1$ from Model 2 at sites i/n $(i=0,\ldots,n)$, for (a)–(c) $n=40$ and $\theta_0=1$, (d)–(f) $n=160$ and $\theta_0=1$, (g)–(i) $n=40$ and $\theta_0=2$, (j)–(l) $n=160$ and $\theta_0=2$.

The distribution of the maximum likelihood estimator of θ_2 is also seriously right-skewed regardless of sample size, and therefore the normal approximations are not appropriate. This is not surprising because $\hat{\theta}_2$ in this case converges again to a non-degenerate and nonnormal variable. It is reasonable to conjecture that the limit of $\hat{\phi}_2$ is also given by (2·7) where $X(s)$ is recovered from $Y(s)$ when $Y(s)$ is observed everywhere on $[0, 1]$. However, it would seem true that the sample size must be quite large in order for the limit distribution to approximate the finite sample distribution well, unless the nugget effect θ_0 is sufficiently small. For all the combinations of sample size and parameter values used here, the approximation given by the limit distribution is rather poor.

5. Discussion

The theoretical and empirical results presented herein highlight the importance of studying asymptotics for spatial statistics, especially infill asymptotics. Indeed, the concept of inconsistent estimability arises only in infill asymptotics because all parameters are consistently estimable under the increasing domain asymptotic framework under some regularity conditions. The fact that infill asymptotics warns that certain covariogram parameters may be hard to estimate and that the estimates may be badly nonnormal even with a large sample size is a compelling virtue of this framework.

In order to make use of existing infill asymptotic results, we restricted attention to one-dimensional stationary Gaussian processes with zero mean, or more generally known mean, and exponential covariogram. Further research is needed to relax some of these restrictions. For example, when the mean is unknown and needs to be estimated, the infill limiting distribution of the maximum likelihood estimator of θ_2 will probably change for the two models in § 3, and one could in this case also consider the infill limit distribution of the restricted maximum likelihood estimator of θ_2. Extensions to higher dimensions and to covariograms other than the exponential are also of interest.

Acknowledgement

The authors thank the associate editor and a referee for helpful comments and suggestions that greatly improved the paper. The research of H. Zhang was partially carried out during visits to the Department of Statistics of the University of British Columbia and to the Department of Statistics and Applied Probability of the National University of Singapore while on sabbatical leave. He thanks Professors Antony Kuk, Wei-Liem Loh, Will Welch and Jim Zidek for support and warm hospitality, and also acknowledges the support of a National Science Foundation grant.

Appendix

Technical Details

Evaluation of (3·8) *for Model* 2. Let

$$a = v_1/(v_0\rho), \quad \Delta = -\{(\rho + \rho^{-1} + a)^2 - 4\}^{0\cdot5}, \quad z_1 = (\rho + \rho^{-1} + a + \Delta)/2.$$

We obtain the following expressions for E_{ij} $(j \geqslant i = 0, 1, 2)$:

$$E_{00} = \frac{1}{2v_0^2}\{1 + 2a(a+\Delta)\Delta^{-2} - a^2(\Delta + 2z_1)\Delta^{-3}\}, \tag{A·1}$$

$$E_{01} = -\frac{1}{2v_0^2\rho}\left(\frac{a+\Delta}{\Delta^2} - \frac{2az_1}{\Delta^3}\right), \tag{A·2}$$

$$E_{02} = \frac{a\delta(1-\rho^2)}{2v_0\rho}(1/\Delta^2 - 2z_1/\Delta^3), \tag{A·3}$$

$$E_{11} = \frac{1}{2v_0^2\rho^2}(\Delta^{-2} - 2z_1\Delta^{-3}), \tag{A·4}$$

$$E_{12} = \frac{\delta}{2a\rho v_0} + \frac{\delta(1-\rho^2)}{2v_0\rho^2}\{1/(a\Delta) + 2z/\Delta^3 - 1/\Delta^2\}, \tag{A·5}$$

$$E_{22} = \frac{\delta^2\rho^2}{1-\rho^2} + \frac{\delta^2}{2} - \frac{\delta^3(1-\rho^2)}{a\rho} + \frac{\delta^2(1-\rho^2)^2}{2\rho^2}\left(\frac{1}{a\Delta^2} - \frac{2}{a\Delta} - \frac{2z_1}{\Delta^2}\right). \quad (A\cdot6)$$

To derive these expressions, first define the complex functions

$$G(z) = z^2 - (\rho + 1/\rho)z + 1, \quad H(z) = z^2 - (\rho + 1/\rho + a)z + 1.$$

Then simple calculations yield

$$1 + \rho^2 - 2\rho\cos\lambda = -\rho e^{-i\lambda}G(e^{i\lambda}), \quad f(\lambda) = \frac{v_0 H(e^{i\lambda})}{2\pi G(e^{i\lambda})}.$$

Using (3·6), we obtain

$$\frac{\partial \log f(\lambda)}{\partial v_0} = \frac{G(e^{i\lambda})}{v_0 H(e^{i\lambda})}, \quad \frac{\partial \log f(\lambda)}{\partial v_1} = -\frac{e^{i\lambda}}{v_0 \rho H(e^{i\lambda})},$$

$$\frac{\partial \log f(\lambda)}{\partial \theta_2} = \delta\rho^{-1}(1-\rho^2)a\frac{e^{2i\lambda}}{G(e^{i\lambda})H(e^{i\lambda})}.$$

The integral in (3·8) can therefore be expressed as a complex integral of rational functions of z on the unit circle $C = \{z : |z| = 1\}$. For example,

$$\int_0^{2\pi}\left\{\frac{\partial \log f(\lambda)}{\partial v_0}\right\}^2 d\lambda = \frac{1}{v_0^2}\int_0^{2\pi}\frac{G^2(e^{i\lambda})}{H^2(e^{i\lambda})}d\lambda = \frac{1}{v_0^2 i}\int_C \frac{G^2(z)}{zH^2(z)}dz. \quad (A\cdot7)$$

Since z_1 is a root of $H(z)$ inside the unit circle, the function $G^2(z)/\{zH^2(z)\}$ is analytic everywhere inside C except at 0 and z_1. The residuals at the points are

$$\operatorname{Res}\left(\frac{G^2(z)}{zH^2(z)}, 0\right) = \frac{G^2(0)}{H^2(0)} = 1, \quad \operatorname{Res}\left(\frac{G^2(z)}{zH^2(z)}, z_1\right) = \lim_{z\to z_1}\frac{\partial}{\partial z}\left\{\frac{G^2(z)}{z(z-z_2)^2}\right\},$$

where $z_2 = (\rho + 1/\rho + a - \Delta)/2$ is the root of H outside the unit circle. After some calculations, we obtain the limit above as $2a(a+\Delta)\Delta^{-2} - a^2(\Delta + 2z_1)\Delta^{-3}$. Applying the Residual Theorem (Rubin, 1966, p. 260), we see that (A·7) equals

$$\frac{1}{v_0^2 i}(2\pi i)\left\{\operatorname{Res}\left(\frac{G^2(z)}{zH^2(z)}, 0\right) + \operatorname{Res}\left(\frac{G^2(z)}{zH^2(z)}, z_1\right)\right\} = (2\pi v_0^{-2})\{1 + 2a(a+\Delta)\Delta^{-2} - a^2(\Delta + 2z_1)\Delta^{-3}\}.$$

Therefore,

$$E_{00} = \frac{1}{4\pi}\int_0^{2\pi}\left\{\frac{\partial \log f(\lambda)}{\partial v_0}\right\}^2 d\lambda = (2v_0^2)^{-1}\{1 + 2a(a+\Delta)\Delta^{-2} - a^2(\Delta + 2z_1)\Delta^{-3}\}.$$

We have derived (A·1). Equations (A·2)–(A·6) can be derived similarly.

Proof of (3·10). Note that $\delta = 1/n$ and ρ, a, Δ and z_1 all depend on n. For two sequences a_n and b_n, we use the notation $a_n \sim b_n$ to mean $\lim_{n\to\infty} a_n/b_n = 1$. For simplicity, we suppress n in all this notation. Then

$$a \sim 2\theta_1\theta_2\delta/\theta_0, \quad \Delta^2 \sim 8\theta_1\theta_2\delta/\theta_0, \quad \rho \sim 1, \quad z_1 \sim 1.$$

Simple calculations yield the following results:

$$E_{00} \sim \frac{1}{2\theta_0^2}, \quad E_{01} \sim \frac{2\theta_1\theta_2}{\theta_0^{3/2}(8\theta_1\theta_2)^{3/2}}\delta^{-1/2}, \quad E_{02} = O(\delta^{3/2}),$$

$$E_{11} \sim \frac{1}{\theta_0^{1/2}(8\theta_1\theta_2)^{3/2}\delta^{3/2}}, \quad E_{12} \sim 1/(4\theta_1\theta_2), \quad E_{22} \sim \delta/(2\theta_2).$$

It follows that $\det(E)$ is dominated by $E_{11}(E_{00}E_{22}-E_{02}^2)\sim E_{11}E_{22}E_{00}$. Hence

$$\det E \sim \{4\theta_0^{5/2}\theta_2(8\theta_1\theta_2)^{3/2}\}^{-1}\delta^{-0\cdot 5}.$$

We are now ready to approximate the elements of E^{-1}. Denote the elements of E^{-1} by Q_{ij} $(i,j=0,1,2)$. Then

$$Q_{00}=(E_{11}E_{22}-E_{12}^2)/\det(E)\sim 2\theta_0^2,\quad Q_{11}=(E_{00}E_{22}-E_{02}^2)/\det(E)\sim \theta_0^{1/2}(8\theta_1\theta_2)^{3/2}\delta^{3/2},$$
$$Q_{22}=(E_{00}E_{11}-E_{01}^2)/\det(E)\sim 2\theta_2/\delta.$$

In addition,

$$Q_{01}\sim -4\theta_0\theta_1\theta_2\delta,\quad Q_{02}=O(\delta^{1/2}),\quad Q_{12}=O(\delta^2).$$

Next, we also approximate the Jacobian matrix J. Let $\phi_1=\theta_1\theta_2=v_1v_2/(1-\rho^2)$. Then

$$\frac{\partial\phi_1}{\partial v_1}=v_2/(1-\rho^2)\sim 1/(2\delta),\quad \frac{\partial\phi_1}{\partial v_2}=v_1\left\{\frac{1}{1-\rho^2}-\frac{2v_2\rho^2\delta}{(1-\rho^2)^2}\right\}\sim 2\theta_1\theta_2\delta.$$

Consequently

$$\left(0,\frac{\partial\phi_1}{\partial v_1},\frac{\partial\phi_1}{\partial v_2}\right)E^{-1}\left(0,\frac{\partial\phi_1}{\partial v_1},\frac{\partial\phi_1}{\partial v_2}\right)'\sim\left(\frac{\partial\phi_1}{\partial v_1}\right)^2 Q_{11}\sim 4(2\theta_0)^{1/2}\phi_1^{3/2}\delta^{-1/2},$$
$$(1,0,0)E^{-1}\left(0,\frac{\partial\phi_1}{\partial v_1},\frac{\partial\phi_1}{\partial v_2}\right)'=Q_{01}\frac{\partial\phi_1}{\partial v_1}+Q_{0,2}\frac{\partial\phi_1}{\partial v_2}\sim -2\theta_0\theta_1\theta_2,$$

and (3·10) follows.

REFERENCES

ABT, M. & WELCH, W. J. (1998). Fisher information and maximum-likelihood estimation of covariance parameters in Gaussian stochastic processes. *Can. J. Statist.* **26**, 127–37.

CHEN, H.-S., SIMPSON, D. G. & YING, Z. (2000). Infill asymptotics for a stochastic process model with measurement error. *Statist. Sinica* **10**, 141–56.

CROWDER, M. J. (1976). Maximum likelihood estimation for dependent observations. *J. R. Statist. Soc.* B **38**, 45–53.

GIHMAN, I. & SKOROHOD, A. V. (1974). *The Theory of Stochastic Processes I*. New York: Springer-Verlag.

GINGRAS, D. F. & MASRY, E. (1988). Autoregressive spectral estimation in additive noise. *IEEE Trans. Acoust. Speech Sig. Proces.* **36**, 490–501.

HALL, P. & HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. New York: Academic Press, Inc.

KARATZAS, I. & SHREVE, S. S. (1991). *Brownian Motion and Stochastic Calculus,* 2nd ed. New York: Springer-Verlag.

LIPTSER, R. S. & SHIRYAYEV, A. A. (1977). *Statistics of Random Processes, I*. New York: Springer-Verlag.

MARDIA, K. V. & MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial statistics. *Biometrika* **71**, 135–46.

PAGANO, M. (1974). Estimation of models of autoregressive signal plus white noise. *Ann. Statist.* **2**, 99–108.

ROSENBLATT, M. (1985). *Stationary Sequences and Random Fields*. Boston: Birkhauser.

RUBIN, W. (1966). *Real and Complex Analysis*. New York: McGraw-Hill Book Company.

STEIN, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging*. New York: Springer-Verlag.

YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *J. Mult. Anal.* **36**, 280–96.

YING, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21**, 1567–90.

ZHANG, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *J. Am. Statist. Assoc.* **99**, 250–61.

[*Received October* 2004. *Revised July* 2005]