PREAMBLE

'Tis a lesson you should heed:
Try, try, try again.
If at first you don't succeed,
Try, try, try again.

W. E. Hickson, **Try and Try Again**

In 1980 Griffith proposed the notion that the Jacobian term for spatial autoregressive models converges upon some constant as the sample size increases to infinity. Several subsequently published pieces severely criticized this idea, using very cogent arguments. Griffith's intuition led him to numerical investigations concerning this issue, and in 1988 he reported convincing but not totally conclusive results supporting it. The purpose of this paper is to report quite conclusive numerical results obtained from supercomputer experiments. Given this brief history, the author is truly optimistic about findings contained in this technical report, even though he constantly bumps into scholars who do not share his enthusiasm (for example, Martin's contribution to this volume). Some of both this enthusiastic and this disheartening viewpoint may be sensed in Ord's commentary on the paper, as he raises questions concerning the sensibleness of answers, on the one hand, and computational manageability, on the other hand.

The Editor

A Numerical Simplification for Estimating Parameters of Spatial Autoregressive Models

Daniel A. Griffith *

Department of Geography, Syracuse University, Syracuse, NY, 13244-1160, U. S. A.

Overview: The Jacobian term appears in likelihood functions to ensure that the use of variable transformations still leads to probability density functions whose complete integration yields unity. This term is particularly troublesome when dealing with spatial autoregressive models, since it does not disappear in the optimization process, and hence requires numerically intensive solutions to the parameter estimation problem. For these sorts of autoregressive models the Jacobian term is a function of the eigenvalues of the connectivity matrix that depicts the geographic configuration of those areal units under study. For a tessellation of n areal units, then, the eigenvalues of an n–by–n matrix need to be calculated. Ord has stated the equations for these eigenvalues when a regular lattice configuration is superimposed upon an infinite surface. Griffith has shown what the algebraic expression of the Jacobian term converges to for this same infinite surface situation. The problem addressed in this paper asks what implications these two simplifications have on parameter estimation for geographically referenced data.

1. Introduction

One reason spatial regression accommodating geographic dependence is so numerically intensive is that the Jacobian of the transformation from an autocorrelated space to an unautocorrelated space must be included in parameter estimation procedures. A Jacobian term is some function of the number of areal units as well as the degree of spatial dependence. Ord (1975) states that the eigenvalues of a binary configuration matrix for a regular lattice are given by the equation

$$\lambda_{kl} = 2\{\cos[k\pi/(n+1)] + \cos[l\pi/(n+1)]\} \tag{1.1}$$

More specifically, the spatial autoregressive parameter is a function of the geographic configuration characterized by this Jacobian, which in most popular models is written in terms of the eigenvalues of the n^2 matrix for this transformation determinant. Questions concerning the accuracy and feasibility of numerically extracted eigenvalues for a given Jacobian, derived from matrices of such large dimensions, have been posed by spatial analysts, and apparently represents a barrier to the dissemination of spatial statistics and spatial econometrics. A frequency distribution for geographic data set size would be sinusoidal or reverse–J shaped; there are numerous data sets where n is quite small, virtually none where n is of moderate

* This research was supported by NSF grant SES–87–22086, with much of its computing being cosponsored by an NSF grant to the National Center for Supercomputing Applications (NCSA), University of Illinois at Urbana–Champaign, for the 1989 Supercomputer Education Summer Workshop, July 10-21. This paper was presented to the Sixth European Colloquium on Theoretical and Quantitative Geography, Chantilly, France, September 5–9, 1989.

size, and some where n is extremely large (most of which are generated from remotely sensed satellite images). The purpose of this paper is to explore ways of mathematically simplifying the calculation of this Jacobian term, especially to help in the analysis of intermediate and large size data sets, and is part of a comprehensive attempt to remove obstacles hindering the diffusion of spatial statistical technology.

1.1. Background

In some respects this present investigation is an extension of two previous undertakings. Griffith (1988a) began exploring possible mathematical simplifications of the Jacobian term for a simultaneous autoregressive model that is based upon a binary connectivity matrix and a regular square lattice configuration of areal units; his study was aimed at remotely sensed data situations. His findings included (1) that there are certain systematic regularities in the Jacobian term as n increases, and (2) even using the analytical equations to compute eigenvalues resulted in considerable rounding error for $10,000$–by–$10,000$ lattices using double–precision FORTRAN on a DEC VAX mainframe. Based upon the numerical findings that were tabulated and reported by Griffith (1988a), searches for the analytical expression describing convergence of the Jacobian term could be restricted to three possible candidates; however, the serious rounding error that prevailed prevented identification of the correct expression from these three.

More recently Griffith (1990) has examined the computation of the Jacobian term, in some cases using double precision, on a Cray 2 supercomputer. His reported results show conclusively that (1) the eigenvalues of a matrix can be computed with a high degree of accuracy for at least $n = 100,000,000$ (this is equivalent to a $10,000$–by–$10,000$ regular square lattice), (2) for a regular square lattice, as n goes to infinity, the Jacobian term for a conditional autoregressive model (which will be the subject of this paper) converges upon the expression

$$-\int_0^{\pi}\int_0^{\pi} \ln\{1 - 2\rho[\cos(\theta_1) + \cos(\theta_2)]\}/\pi^2 \; d\theta_1 \, d\theta_2, \tag{1.2}$$

which can be numerically integrated (see Table 1), and (3) that the parameter estimation impact of the Jacobian term diminishes in importance as n goes to infinity. Equation (1.2) is the continuous version of and is converged upon by

$$-\sum_{k=1}^{m}\sum_{l=1}^{m} \ln\Big[1 - 2\rho\{\cos[k\pi/(m+1)] + \cos[l\pi/(m+1)]\}\Big]/n, \tag{1.3}$$

where $m^2 = n$ (or $m = \sqrt{n}$), which represents the Jacobian term for finite square lattices. As one can see from expressions (1.2) and (1.3), the Jacobian term is a mean; it should not be surprising from a statistical perspective, then, to find that this quantity converges as n goes to infinity.

These two previous studies have set the stage for the analysis presented in this paper. Here attention will be restricted to the Jacobian term for a conditional autoregressive model.

TABLE 1
SELECTED RESULTS FOR THE NUMERICAL SOLUTION OF EQUATION (1.2)

rho	integral value	error	rho	integral value	error
0.025	0.0012535321	0.0000000000	0.150	0.0505218648	0.0000000000
0.050	0.0050573165	0.0000000000	0.200	0.1014553111	0.0000000000
0.100	0.0209735079	0.0000000000	0.250	0.2200507460	0.0000000003

NOTE: Numerical integration has been achieved with the IMSL10 routine E2LSF.

2. A Jacobian term equation with ρ varying for selected n

Two interesting limiting cases of n for which one might assess variation in the Jacobian term as the spatial autoregressive parameter ρ changes are its lower limit, where for a regular square lattice $\sqrt{n} = 2$, and its upper limit, where for a regular square lattice $\sqrt{n} = \infty$. Five–hundred–and–one Jacobian terms were computed for each of these two cases, using values of ρ that started with the limiting parameter space boundary -0.25, and were sequentially incremented by 0.001, until 0.25 was reached. An analysis of these two sets of results lead to the formulation of an equation describing how the Jacobian term changes over the possible natures and degrees of spatial dependence. Next, various intermediate values of $\sqrt{n}$ were studied, using twenty–one uniformly spaced values of ρ across the feasible parameter space (namely, ± 0.25, ± 0.225, ± 0.2, ± 0.175, ± 0.15, ± 0.125, ± 0.10, ± 0.075, ± 0.5, ± 0.025, and 0.0), and yielded the tabulated numerical results presented in Table 2. General tendencies present in this table include (1) a mean squared error value that increases with n, but never to a non–negligible level, (2) asymptotically converging estimates for the two parameters β_n and γ_n, with very little difference in subsequent values for these parameters beyond $n = 900$ (the correlation between results for $n = 2^2$ and $n = \infty$ is 0.987), and (3) a value of β_n which is approximately twice the value of γ_n. To illustrate these findings within their equational context, the two limiting cases would yield the following equations:

$$\sqrt{n} = 2: \quad J = \ln(0.5)/2 - 0.25\,\ln(0.5+\rho) - 0.25\,\ln(0.5-\rho), \text{ and}$$
$$\sqrt{n} = \infty: \; J = -0.377580 - 0.150659\,\ln(0.285620+\rho) - 0.150659\,\ln(0.285620-\rho).$$

These findings imply that the Jacobian term is a concave–upwards function, whose general form is

$$J = 2\beta_n \ln(\gamma_n) - \beta_n \ln(\gamma_n + \rho) - \beta_n \ln(\gamma_n - \rho), \tag{2.1}$$

having an increasingly shallower trough as n increases. This feature is consistent with the aforementioned contention that the importance of the Jacobian term diminishes as n increases. As is indicated by their subscripts, the parameters β_n and γ_n are functions of the number of areal units under study.

As $\sqrt{n}$ increases toward infinity, a slight bias seems to appear in the computations of β_n and γ_n. This slight bias may well be attributable to the residual heteroscedasticity stemming from some systematic error component arising in the numerical eigenvalue extraction algorithm, or possibly from specification error.

Presumably for a regular rectangular lattice equation (2.1) would become

$$J = \alpha_n - \beta_n \ln(\gamma_n + \rho) - \beta_n \ln(\gamma_n - \rho), \tag{2.2}$$

TABLE 2
NONLINEAR REGRESSION PARAMETER ESTIMATES
OF SELECTED JACOBIAN TERMS

$\sqrt{n}$ of square lattice	β_n	γ_n	MSE	$\sqrt{n}$ of square lattice	β_n	γ_n	MSE
2	0.250000	0.500000	0.000000000	54	0.153596	0.289884	0.000001983
4	0.185791	0.354859	0.000000029	56	0.153498	0.289733	0.000002005
6	0.173667	0.327639	0.000000155	58	0.153405	0.289592	0.000002026
8	0.168056	0.315926	0.000000331	60	0.153318	0.289461	0.000002045
10	0.164745	0.309377	0.000000512	62	0.153237	0.289338	0.000002064
12	0.162536	0.305183	0.000000680	64	0.153162	0.289224	0.000002081
14	0.160951	0.302264	0.000000830	66	0.153091	0.289116	0.000002097
16	0.159752	0.300112	0.000000963	68	0.153023	0.289014	0.000002112
18	0.158816	0.298462	0.000001080	70	0.152958	0.288917	0.000002127
20	0.158058	0.297151	0.000001183	72	0.152900	0.288828	0.000002142
22	0.157433	0.296087	0.000001275	74	0.152842	0.288741	0.000002155
24	0.156909	0.295205	0.000001357	76	0.152787	0.288660	0.000002168
26	0.156462	0.294461	0.000001430	78	0.152736	0.288583	0.000002179
28	0.156077	0.293826	0.000001495	80	0.152687	0.288510	0.000002191
30	0.155742	0.293278	0.000001555	82	0.152640	0.288440	0.000002202
32	0.155446	0.292798	0.000001609	84	0.152594	0.288373	0.000002213
34	0.155184	0.292376	0.000001658	86	0.152551	0.288310	0.000002223
36	0.154950	0.292002	0.000001704	88	0.152512	0.288250	0.000002233
38	0.154737	0.291666	0.000001745	90	0.152470	0.288191	0.000002243
40	0.154547	0.291365	0.000001783	92	0.152434	0.288136	0.000002252
42	0.154375	0.291093	0.000001818	94	0.152397	0.288083	0.000002260
44	0.154217	0.290845	0.000001850	96	0.152363	0.288033	0.000002268
46	0.154072	0.290620	0.000001881	98	0.152330	0.287984	0.000002277
48	0.153940	0.290413	0.000001909	100	0.152298	0.287937	0.000002284
50	0.153816	0.290222	0.000001935	1000	0.150835	0.285857	0.000002664
52	0.153703	0.290047	0.000001960	∞	0.150659	0.285620	0.000002712

NOTE 1: The case of infinity had the Jacobian terms computed with IMSL10 subroutine E2LSF (see Table 1).

NOTE 2: Using a division of the feasible parameter space $[-0.25, 0.25]$ into 501 values resulted in almost exactly the same values for the case of $\sqrt{n} = 2$, but noticeably different values for $\sqrt{n} = \infty$ ($\beta_\infty = 0.163846$, $\gamma_\infty = 0.293988$, and MSE $= 0.000001457$). This latter inconsistency suggests that either the numerical integration, or the numerical eigenvalue extraction, is plagued with error.

NOTE 3: For the above cases $R^2 = 1.000$, the Wilk–Shapiro statistic $= 0.969$ for $\sqrt{n} = 2$ and asymptotically converges on 0.931 as n increases, the Durbin–Watson statistic $= 2.82$ for $\sqrt{n} = 2$ and apparently converges in an oscillatory fashion on 2.10 as n increases, and there is no apparent heteroscedasticity displayed by the residuals for $\sqrt{n} = 2$, with increasingly systematic, complex, nonlinear heteroscedasticity displayed by the residuals as n increases, together with extreme values becoming influential estimation points as n increases.

where the parameter α_n no longer is constrained to be a function of β_n and γ_n [such as the term $2\beta_n \ln(\gamma_n)$ appearing in equation (2.1)], and for an irregular lattice equation (2.1)

would become

$$J = \alpha_n - \beta_{1,n}\ln(\gamma_{1,n} + \rho) - \beta_{2,n}\ln(\gamma_{2,n} - \rho). \tag{2.3}$$

These last two conjectures require considerable subsequent investigation.

Given equation (2.1), the log–likelihood function, say $\ln(L)$, to be optimized when calculating a maximum likelihood estimate of the spatial autoregressive parameter ρ becomes

$$\begin{aligned} K - (n/2)\ln(\sigma^2) - (n/2)[2\beta_n\gamma_n - \beta_n\ln(\gamma_n + \rho) - \beta_n\ln(\gamma_n - \rho)] \\ - (\mathbf{X} - \mu\mathbf{1})^t(\mathbf{I} - \rho\mathbf{C})(\mathbf{X} - \mu\mathbf{1})/(2\sigma^2), \end{aligned} \tag{2.4}$$

where $K = -(n/2)\ln(2\pi)$ is a constant, $\mathbf{X}$ is an n–by–1 data vector, $\mathbf{1}$ is an n–by–1 vector of ones, and $\mathbf{C}$ is an n–by–n binary geographic configuration matrix (upon which the Jacobian term is based). Parameter estimation based upon equation (2.4) currently requires nonlinear optimization techniques.

3. Parameter estimation based upon the simplified Jacobian term

Four different estimation cases can be explored for the likelihood function portrayed by expression (2.4), each referring to a combinatorial possibility of unknown parameter values. In the first of these cases, suppose that only ρ is unknown (in other words, let μ and σ^2 be known). Optimizing expression (2.4) with respect to ρ yields

$$\frac{\partial\ln(L)}{\partial\rho} = -(n/2)[-\beta_n/(\gamma_n + \rho) + \beta_n/(\gamma_n - \rho)] + [(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mu\mathbf{1})]/(2\sigma^2) = 0,$$

which when solved produces a quadratic equation in ρ having roots

$$\hat{\rho} = -n\sigma^2\beta_n/(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mu\mathbf{1}) \pm \{[-n\sigma^2\beta_n/(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mu\mathbf{1})]^2 + \gamma_n^2\}^{1/2}. \tag{3.1}$$

Consequently, the spatial autocorrelation parameter becomes an explicit function of the size of the geographic data series, as well as the configuration of the underlying areal unit surface partitioning. One should expect this definition of ρ always to be real, and always to fall within the feasible parameter space region. This finding is particularly useful for remotely sensed data analysis, for once the parameters β_n and γ_n are established, then all one needs to know is the size of the regular square lattice partitioning in order to estimate ρ; numerical computation of eigenvalues no longer will be necessary.

In the second case one can assume that both μ and ρ are unknown, and only σ^2 is known. Now optimizing expression (2.4) with respect to μ and ρ yields the standard maximum likelihood estimation (MLE) result of

$$\hat{\mu} = \mathbf{1}^t(\mathbf{I} - \rho\mathbf{C})\mathbf{X}/\mathbf{1}^t(\mathbf{I} - \rho\mathbf{C})\mathbf{1}, \tag{3.2}$$

and hence the differential equation

$$\frac{\partial\ln(L)}{\partial\rho} = -(n/2)[-\beta_n/(\gamma_n + \rho) + \beta_n/(\gamma_n - \rho)] + [(\mathbf{X} - \hat{\mu}\mathbf{1})^t\mathbf{C}(\mathbf{X} - \hat{\mu}\mathbf{1})]/(2\sigma^2) = 0,$$

which when solved produces a quartic equation in ρ of the form

$$
\begin{aligned}
&-[(1^t\mathbf{C}1\mathbf{X}^t - 1^t\mathbf{C}\mathbf{X}1^t)\mathbf{C}\mathbf{X}1^t\mathbf{C}1]\rho^4 \\
&+2[-n\beta_n\sigma^2(1^t\mathbf{C}1)^2 + n\mathbf{X}^t\mathbf{C}\mathbf{X}1^t\mathbf{C}1 - n(1^t\mathbf{C}\mathbf{X})^2]\rho^3 \\
&+\{(1^t\mathbf{C}1\mathbf{X}^t - 1^t\mathbf{C}\mathbf{X}1^t)\mathbf{C}\mathbf{X}1^t\mathbf{C}1\gamma_n^2 \\
&\quad +[4n^2\beta_n\sigma^2 - (1^t\mathbf{X})^2]1^t\mathbf{C}1 - n^2\mathbf{X}^t\mathbf{C}\mathbf{X} + 2n1^t\mathbf{C}\mathbf{X}1^t\mathbf{X}\}\rho^2 \\
&-2\{n^3\beta_n\sigma^2 + n[\mathbf{X}^t\mathbf{C}\mathbf{X}1^t\mathbf{C}1 - (1^t\mathbf{C}\mathbf{X})^2]\gamma_n^2\}\rho \\
&+[n^2\mathbf{X}^t\mathbf{C}\mathbf{X} + (1^t\mathbf{C}11^t\mathbf{X} - 2n1^t\mathbf{C}\mathbf{X})1^t\mathbf{X}]\gamma_n^2 = 0.
\end{aligned}
$$

The solution to a biquadratic equation of this kind is presented in theory of equation texts, such as the classic by Uspensky (1948, pp. 94-97). An algorithmic solution to solving the underlying pair of simultaneous differential equations also can be pursued, if one wishes to avoid extracting roots of a fourth–order equation. The iterative algorithm would be of the form

Step 1: let $\rho = 0 \Rightarrow \hat{\mu} = \mathbf{1}^t\mathbf{X}/n$ (for iteration $\tau = 0$);

Step 2: solve $\hat{\rho}$ for equation (3.1), in Case I;

Step 3: compute $\hat{\mu}_{\tau+1} = \mathbf{1}^t(\mathbf{I} - \hat{\rho}_\tau\mathbf{C})\mathbf{X}/\mathbf{1}^t(\mathbf{I} - \hat{\rho}_\tau\mathbf{C})\mathbf{1}$; and,

Step 4: iterate through Steps 2 and 3 until the parameter estimates converge (this and subsequent algorithms are believed to converge, although no proof of convergence is offered here; at worst they should be good heuristics).

For the third case consider both σ^2 and ρ to be unknown, with only μ being known (for example, the case of regression residuals). Now optimizing expression (2.4) with respect to σ^2 and ρ yields the standard MLE result of

$$\hat{\sigma}^2 = (\mathbf{X} - \mu\mathbf{1})^t(\mathbf{I} - \rho\mathbf{C})(\mathbf{X} - \mu\mathbf{1})/n, \tag{3.3}$$

and hence the differential equation

$$\frac{\partial \ln(L)}{\partial \rho} = -(n/2)[-\beta_n/(\gamma_n + \rho) + \beta_n/(\gamma_n - \rho)] + [(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mu\mathbf{1})]/(2\hat{\sigma}^2) = 0,$$

which when solved produces a quadratic equation in ρ having roots

$$
\begin{aligned}
\hat{\rho} = &- [\beta_n/(1 - 2\beta_n)][(\mathbf{X} - \mu\mathbf{1})^t(\mathbf{X} - \mu\mathbf{1})/(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mu\mathbf{1})] \\
&\pm \{[\beta_n^2/(1 - 2\beta_n)^2][(\mathbf{X} - \mu\mathbf{1})^t(\mathbf{X} - \mu\mathbf{1})/(\mathbf{X} - \mu\mathbf{1})^t\mathbf{C}(\mathbf{X} - \mathbf{1})]^2 \\
&+ \gamma_n^2/(1 - 2\beta_n)\}^{1/2}.
\end{aligned} \tag{3.4}
$$

An algorithmic solution to solving the underlying pair of simultaneous differential equations can be pursued in this case, too, although it is doubtful if one ever would seriously wish to avoid calculating the pair of roots. The iterative algorithm would be of the form listed in the following steps:

1: let $\rho = 0 \Rightarrow \hat{\sigma}^2 = (\mathbf{X} - \mu\mathbf{1})^t(\mathbf{X} - \mu\mathbf{1})/n$ (for iteration $\tau = 0$);

2: solve $\hat{\rho}$ for equation (3.1), in Case I;

3: compute $\hat{\sigma}^2_{\tau+1} = (\mathbf{X} - \mu\mathbf{1})^t(\mathbf{I} - \hat{\rho}_\tau\mathbf{C})(\mathbf{X} - \mu\mathbf{1})/n$; and,

4: iterate through Steps 2 and 3 until the parameter estimates converge.

The fourth, and final, case to be treated here is the more likely situation that none of the parameters are known, or μ, σ^2 and ρ are unknown. Here optimizing expression (2.4) with respect to μ, σ^2 and ρ yields the standard MLE results appearing in equations (3.2) and (3.3), as well as the differential equation

$$\frac{\partial \ln(L)}{\partial \rho} = -(n/2)[-\beta_n/(\gamma_n+\rho) + \beta_n/(\gamma_n-\rho)] + [(\mathbf{X}-\hat{\mu}\mathbf{1})^{\mathbf{t}}\mathbf{C}(\mathbf{X}-\hat{\mu}\mathbf{1})]/(2\sigma^2) = 0,$$

which when solved produces a quartic equation in ρ of the form

$$\begin{aligned}
&(2\beta_n - 1)[(\mathbf{1}^t\mathbf{C}\mathbf{1}\mathbf{X}^t - \mathbf{1}^t\mathbf{C}\mathbf{X}\mathbf{1}^t)\mathbf{C}\mathbf{X}\mathbf{1}^t\mathbf{C}\mathbf{1}]\rho^4 \\
&+ 2\{(2\beta_n - 1)n[(\mathbf{1}^t\mathbf{C}\mathbf{X})^2 - \mathbf{X}^t\mathbf{C}\mathbf{X}\mathbf{1}^t\mathbf{C}\mathbf{1}] \\
&\quad - \beta_n[(\mathbf{1}^t\mathbf{C}\mathbf{1})^2\mathbf{X}^t\mathbf{X} - 2\mathbf{1}^t\mathbf{C}\mathbf{1}\mathbf{1}^t\mathbf{X}\mathbf{1}^t\mathbf{C}\mathbf{X} + n(\mathbf{1}^t\mathbf{C}\mathbf{X})^2]\}\rho^3 \\
&+ \{4\beta_n\mathbf{1}^t\mathbf{C}\mathbf{1}[n\mathbf{X}^t\mathbf{X} - (\mathbf{1}^t\mathbf{X})^2] + \gamma_n^2\mathbf{1}^t\mathbf{C}\mathbf{1}[\mathbf{X}^t\mathbf{C}\mathbf{X}\mathbf{1}^t\mathbf{C}\mathbf{1} - (\mathbf{1}^t\mathbf{C}\mathbf{X})^2] \\
&\quad + (2\beta_n - 1)[n^2\mathbf{X}^t\mathbf{C}\mathbf{X} - 2n\mathbf{1}^t\mathbf{C}\mathbf{X}\mathbf{1}^t\mathbf{X} + \mathbf{1}^t\mathbf{C}\mathbf{1}(\mathbf{1}^t\mathbf{X})^2]\}\rho^2 \\
&+ 2n\{\beta_n[(\mathbf{1}^t\mathbf{X})^2 - n\mathbf{X}^t\mathbf{X}] + \gamma_n^2[(\mathbf{1}^t\mathbf{C}\mathbf{X})^2 - \mathbf{1}^t\mathbf{C}\mathbf{1}\mathbf{X}^t\mathbf{C}\mathbf{X}]\}\rho \\
&+ [n^2\mathbf{X}^t\mathbf{C}\mathbf{X} + (\mathbf{1}^t\mathbf{C}\mathbf{1}\mathbf{1}^t\mathbf{X} - 2n\mathbf{1}^t\mathbf{C}\mathbf{X})\mathbf{1}^t\mathbf{X}]\gamma_n^2 = 0. \qquad (3.5)
\end{aligned}$$

Two possible algorithmic solutions to solving the underlying triplet of simultaneous differential equations can be pursued in this case, if one wishes to avoid calculating the roots of a fourth–degree polynomial. One iterative algorithm could be of the form listed in the following steps:

A–1: let $\rho = 0 \Rightarrow \hat{\mu} = \mathbf{1}^t\mathbf{X}/n$ and $\hat{\sigma}^2 = (\mathbf{X}-\hat{\mu}\mathbf{1})^t(\mathbf{X}-\hat{\mu}\mathbf{1})/n$ (for iteration $\tau = 0$);

A–2: solve $\hat{\rho}$ for equation (3.1), in Case I;

A–3: first compute

$$\hat{\mu}_{\tau+1} = \mathbf{1}^t(\mathbf{I}-\hat{\rho}_\tau\mathbf{C})\mathbf{X}/\mathbf{1}^t(\mathbf{I}-\hat{\rho}_\tau\mathbf{C})\mathbf{1},$$

and then compute

$$\hat{\sigma}^2_{\tau+1} = (\mathbf{I}-\hat{\mu}_{\tau+1}\mathbf{1})^t(\mathbf{I}-\hat{\rho}_\tau\mathbf{C})(\mathbf{X}-\hat{\mu}_{\tau+1}\mathbf{1})/n;$$

and,

A–4: iterate through Steps A–2 and A–3 until the parameter estimates converge.

An alternative algorithm would be of the form listed in the following steps:

B–1: let $\rho = 0 \Rightarrow \hat{\mu} = \mathbf{1}^t\mathbf{X}/n$ (for iteration $\tau = 0$);

B–2: solve $\hat{\rho}$ for equation (3.4), in Case III;

B–3: compute $\hat{\mu}_{\tau+1} = \mathbf{1}^t(\mathbf{I}-\hat{\rho}_\tau\mathbf{C})\mathbf{X}/\mathbf{1}^t(\mathbf{I}-\hat{\rho}_\tau\mathbf{C})\mathbf{1}$;

B–4: iterate through Steps B–2 and B–3 until the parameter estimates converge; and,

B–5: $\hat{\sigma}^2 = (\mathbf{I}-\hat{\mu}\mathbf{1})^t(\mathbf{I}-\hat{\rho}\mathbf{C})(\mathbf{X}-\hat{\mu}\mathbf{1})/n$.

Intuitively speaking, for both Cases II and IV, at least two of the roots of their fourth–degree equations [such as (3.5)] must be real; accordingly, at most two roots can be complex.

An interesting mathematical exercise would be to prove this conjecture, perhaps using Ferrari's solution technique for biquadratic equations. At most, one of the roots should fall into the feasible parameter space; conditions governing the existence of this category of solution need to be established. The remaining four combinatorial possibilities for known and unknown parameters are of no interest here, since they do not involve the estimation of the spatial autoregressive parameter ρ. Finally, if the expression $(\mathbf{I} - \hat{\mu}\mathbf{1})^t\mathbf{C}(\mathbf{X} - \hat{\mu}\mathbf{1}) = 0$, then $\hat{\rho} = 0$.

4. Examples

A contrived example will be described in this section in order to illustrate the four estimation cases outlined in the preceding section. For this example imagine that $\mu = 0$, $\sigma^2 = 1$, $n = 4$ (so $\beta_2 = 0.25$ and $\gamma_2 = 0.5$; see Table 2), and the geographic distribution in question is

0	0
1	-1

for which $\mathbf{1}^t\mathbf{X} = 0$, $\mathbf{X}^t\mathbf{X} = 2$, $\mathbf{1}^t\mathbf{C}\mathbf{X} = 0$, and $\mathbf{X}^t\mathbf{C}\mathbf{X} = -2$. The four eigenvalues of matrix $\mathbf{C}$ for this geographic configuration are -2, 0, 0, and 2; hence, the spatial autoregressive parameter estimate has the restriction $-1/2 < \hat{\rho} < 1/2$.

Case I:

$$\hat{\rho} = 2\beta_n \pm (4\beta_n^2 + \gamma_n^2)^{1/2} = (1 \pm \sqrt{2})/2,$$

which means that $\hat{\rho} = (1 - \sqrt{2})/2$ satisfies the accompanying constraint;

Case II:

$$\begin{aligned} 4\rho^4 - 4(4\beta_n + 1)\rho^3 + (16\beta_n - 4\gamma_n^2 + 1)\rho^2 - 4(\beta_n - \gamma_n^2)\rho - \gamma_n^2 &= 0 \\ (\rho^2 - 4\beta_n\rho - \gamma_n^2)(4\rho^2 - 4\rho + 1) &= 0 \\ (\rho^2 - 4\beta_n\rho - \gamma_n^2) &= 0 \end{aligned}$$

is the equation for Case I, which is what would be expected since the sample mean, the MLE of the mean, and the population mean are identical;

Case III:

$$\begin{aligned} \hat{\rho} &= \beta_n/(1 - 2\beta_n) \pm [\beta_n^2/(1 - 2\beta_n)^2 + \gamma_n^2/(1 - 2\beta_n)]^{1/2} \\ &= (1 \pm \sqrt{3})/2, \end{aligned}$$

which means that $\hat{\rho} = (1 - \sqrt{3})/2$ satisfies the accompanying constraint; and,

Case IV:

$$\begin{aligned} 8(2\beta_n - 1)\rho^4 + 8\rho^3 + 2(4\gamma_n^2 - 6\beta_n - 1)\rho^2 + 4(\beta_n - 2\gamma_n^2)\rho + 2\gamma_n^2 &= 0 \\ [(2\beta_n - 1)\rho^2 + 2\beta_n\rho + \gamma_n^2](8\rho^2 - 8\rho + 2) &= 0 \\ (2\beta_n - 1)\rho^2 + 2\beta_n\rho + \gamma_n^2 &= 0 \end{aligned}$$

is the equation for Case III, which again is what would be expected since the sample mean, the MLE of the mean, and the population mean are identical.

Two interesting observations can be made about these findings. First, Case II will reduce to Case I, and Case IV will reduce to Case III, in selected situations. Second, different estimates of ρ are obtained for different levels of ignorance (just like with the classical sample variance).

5. Simulation experiment results

As the reported error (MSE) in Table 2 indicates, while equation (2.1) furnishes an exact Jacobian term for the regular square lattice situation of $\sqrt{n} = 2$, all other square lattice sizes have some very small (and seemingly negligible) amount of error present. Since the preceding example is based upon this exact situation, a simple simulation experiment has been conducted for $\sqrt{n} = 4$ to explore whether or not the Jacobian approximation approach promoted in this paper accurately generalizes to larger square lattice cases.

The initial conditions of this simulation experiment are (1) 16 values were randomly generated with the MINITAB normal pseudo–random number generator, having $\mu = 0$ and $\sigma^2 = 1$, and (2) $\mathbf{1}^t\mathbf{C}\mathbf{1} = 48$, and $\beta_n = 0.185791$ and $\gamma_n = 0.354859$ (see Table 2). The resulting geographic distribution of generated sample values is

-0.92327	1.31724	-0.99017	1.07651
-0.50171	1.82117	-0.29935	-1.13190
0.28974	-0.98088	-0.52315	-1.10900
1.70435	0.81597	0.28311	-1.47315

The sample statistic terms for this spatial arrangement are $\mathbf{1}^t\mathbf{X} = -0.62449$, $\mathbf{1}^t\mathbf{C}\mathbf{X} = -2.24012$, $\mathbf{X}^t\mathbf{X} = 18.03662$, and $\mathbf{X}^t\mathbf{C}\mathbf{X} = -0.18677$. The sixteen eigenvalues for this geographic configuration are 3.23607, 2.23607, 2.23607, 1.23607, 1.00000, 1.00000, 0.00000, 0.00000, 0.00000, 0.00000, −1.00000, −1.00000, −1.23607, −2.23607, −2.23607, and −3.23607. In addition, the classical statistics for this sample surface are $\bar{\mathbf{x}} = -0.03903$, $s = 1.09582$, and the modified Wilk–Shapiro = 0.96217. The traditional estimation procedure for obtaining $\hat{\rho}$, when all three of the parameters are unknown (Case IV above), solves the following optimization problem (see Upton and Fingleton, 1985):

$$\begin{aligned} MIN : & [\prod_{i=1}^{16}(1-\rho\lambda_i)]^{-1/16}(\mathbf{X}-\mu\mathbf{1})^t(\mathbf{I}-\rho\mathbf{C})(\mathbf{X}-\mu\mathbf{1}) \\ \text{st} : & -1/3.23607 < \hat{\rho} < 1/3.23607, \end{aligned} \tag{5.1}$$

where λ_i $(i = 1, 2, \ldots, 16)$ are the sixteen aforementioned eigenvalues of the binary configuration matrix $\mathbf{C}$. The solution to this particular problem, using the IMSL10 subroutines

E2LSF (to extract eigenvalues) and UVMID (to achieve univariate nonlinear optimization) [in single precision on a VAX mainframe], yielded $\hat{\rho} = -0.00541$, and $\hat{\mu} = -0.03915$.

Appropriate substitutions into equation (3.5) produce the quartic equation

$$421.7902\rho^4 - 15702.7248\rho^3 + 10242.3840\rho^2 - 1657.0680\rho - 9.3008 = 0. \qquad (5.2)$$

The roots of equation (5.2) have been extracted using the IMSL10 subroutine ZPORC (in single precision on a VAX mainframe), and are -0.00543, 0.33189, 0.33467, and 36.56763. Of these four roots, the only one that falls within the feasible parameter space interval $(-0.30930, 0.30930)$ is -0.00543, which is equivalent to that obtained with the nonlinear optimization of expression (5.1), except for rounding error. This illustration demonstrates that, indeed, the set of equations (3.2), (3.4), and (3.5), involving Jacobian term approximations, do render very accurate estimates, and dramatically reduce the numerical intensity of spatial autoregression analysis.

6. Concluding comments and implications

The Jacobian term appears in likelihood functions to ensure that principal components types of transformations still lead to probability density functions whose complete integration yields unity. This term is particularly troublesome when dealing with spatial autoregressive models, since it is a function of the prevailing nature and degree of spatial dependence, becomes complex because of the multi–directional and two–dimensional interdependence involved, and thus does not disappear in the optimization process. Historically this term has required numerically intensive, and perhaps often computationally inaccurate, solutions to the parameter estimation problem. Findings reported in this paper suggest that at least a closed form approximation to this Jacobian may exist. The form of this approximation for rectangular regular lattices, and irregular lattices, still needs to be identified. The accuracy of this approximation remains to be comprehensively studied.

Having an approximation that is relatively simple in form, like equations (3.1) and (3.4), should allow a more careful and clearer investigation of the statistical properties of bias, sufficiency, consistency, and efficiency, for the parameter estimate $\hat{\rho}$. Ord (1975) already has reported some findings pertaining to these characteristics. The formulation presented here also will facilitate a better comparison between ordinary and generalized least squares parameter estimates for spatial regression models. Hopefully the formulations uncovered here will afford deeper insights into these statistical properties. In addition, equations need to be established depicting the convergence, as n increases, of the Jacobian term parameters. Attempts thus far to achieve this goal have failed, but were for the Jacobian terms themselves. Finally, direct extensions to the moving average and simultaneous autoregressive model, as well as to stochastic versions of the geographic configuration matrix $\mathbf{C}$, merit careful attention.

Meanwhile, interfacing these findings with previous projects has some interesting implications. Griffith (1988b, 1989) has made a concerted effort to translate spatial regression techniques into algebraic language that is compatible with commercial statistical software packages. The general approach employed is to attach weights when writing regression equations, in much the same way that weighted least squares regression does. The weight that is attached is a function of the Jacobian term studied in this paper. By being able to write

this term in a simplified and consolidated form, rather than as a sum of eigenvalue expressions, these sorts of efforts with commercial packages will be further enhanced. The principal drawback here, though, is that the theory upon which optimization is based in these reformulated situations may be inapplicable; Warnes and Ripley (1987) have cast some doubt on the soundness of this approach, although they comment primarily on the parametric covariance function rather than an autoregressive formulation. Nevertheless, at least a good and useful first–approximation may be obtainable from these techniques.

7. References

Griffith, D. (1988a) Jacobian term specification and parameter estimation for spatial autoregressive models. Paper presented to the Association of American Geographers, 84th Annual Meeting, Phoenix.

Griffith, D. (1988b) Estimating spatial autoregressive model parameters with commercial statistical packages. *Geographical Analysis*. **20**, 176-186.

Griffith, D. (1989) Spatial regression analysis on the PC: spatial statistics using `MINITAB`. *Discussion Paper #1*. Ann Arbor: Institute of Mathematical Geography.

Griffith, D. (1990) Supercomputing and spatial statistics: a reconnaissance. *The Professional Geographer*. **42**, forthcoming.

Ord, J. (1975) Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*. **70**, 120-126.

Upton, G., and B. Fingleton (1985) *Spatial Data Analysis by Example*. Vol. 1. New York: Wiley.

Uspensky, J. (1948) *Theory of Equations*. New York: McGraw–Hill.

Warnes, J. and B. Ripley (1987) Problems with likelihood estimation of covariance functions of spatial Gaussian processes. *Biometrika*. **74**, 640-642.

DISCUSSION

"A numerical simplification for estimating parameters of spatial autoregressive models"

by Daniel A. Griffith

Estimation of spatial autoregressive (AR) processes has proved to be a very awkward statistical problem. For small or moderate numbers of areal units, the computational aspects are manageable, although the performance of the maximum likelihood procedure may be disappointing, manifested in a wide confidence band or a very flat likelihood function (LF). Worse yet, when a parametric covariance function is used, Warnes and Ripley (1987) show that the LF may have multiple maxima and that these need not relate well to the true parameter values, as is illustrated by simulation results in Ripley (1988, pp. 15-19). The reasons for this are not fully understood, but seem to be bound up with the scale of the process. That is, the AR structure is modeled primarily as a local phenomenon, yet longer range effects may have a major impact on the estimation process.

A second aspect of this is the pattern of sample covariances, first noted in the time series context by Anderson (1981), but likely to apply *a fortiori* in the spatial case. Suppose that there are n observations, yielding $\mathbf{N} = \binom{\mathbf{n}}{\mathbf{2}}$ pairings classified into $\mathbf{K}$ groups. For example, we may group by distances; if $\mathbf{d(i,j)}$ represents the distance between the pair of locations $\mathbf{(i,j)}$, the $\mathbf{k^{th}}$ group may be defined as $\mathbf{d_{k-1} < d(i,j) \leq d_k}, \quad \mathbf{k = 1, \ldots, K}$. Typically, $\mathbf{d_0 = 0}$ and $\mathbf{d_K}$ is the maximum distance between locations in the study area. Define

$$\mathbf{c_k = \sum (x_i - \bar{x})(x_j - \bar{x})/N_k, \quad k = 1, \ldots, K,} \tag{1}$$

where the sum is taken over pairs $\mathbf{(i,j)}$ in class $\mathbf{k}$, and $\mathbf{N_k}$ denotes the number of such pairs. Then it is easily shown that

$$\mathbf{nc_0 + 2 \sum N_k c_k = \sum_i \sum_j (x_i - \bar{x})(x_j - \bar{x}),} \tag{2}$$

where $\mathbf{nc_0 = \sum (x_i - \bar{x})^2}$. But the right hand side of equation (2) is identically zero so that the autocorrelations, $\mathbf{r_k = c_k / c_0}$ satisfy

$$\mathbf{\sum m_k r_k = -1/2,} \tag{3}$$

where $\mathbf{m_k = N_k / n}$. Thus, even when the theoretical ACF is nonnegative, as is often assumed, equation (3) *requires* that some of the sample values are negative. Since $\mathbf{K}$ is of the order of $\mathbf{n^{1/2}}$ for two–dimensional processes (against $\mathbf{K = n - 1}$ for time series), this poses real problems. Combined with the knowledge that ρ rapidly approaches its upper bound as the autocorrelation increases (cf. Bartlett, 1975, pp. 82-83), the conclusions must be that the sample ACF is rather uninformative and the ML estimators may not be reliable either.

For the very large samples often encountered in image processing, the dominant problem is computational, although one must still worry whether the global assumptions of stationarity is justified; weaker assumptions such as the existence of the variogram seem easier to sustain.

Where then do these comments lead us? In many cases, ρ will be close to its upper bound and so an approximation to the Jacobian element of the likelihood that identifies the bound will often lead to an estimate close to the actual MLE. The approximate large sample variance may be inaccurate, but this could be improved by examining approximations to the second derivative of expression (1.2) in Griffith's paper. Thus, for larger samples, the computational burden is greatly eased; for smaller samples, I suspect that when we are close to the boundary of the parameter space, nothing will be of much help.

If nonstationarity is suspected for large samples, the study area may be partitioned and the estimates obtained for each subarea. Griffith's proposals make such exploratory analyses much more accessible. Overall, past results lead us to be cautious in introducing approximate methods, but such an approach may lead to better data analysis.

We now turn to the particular approximations suggested by Griffith. Rather than the computer intensive search process suggested above his equation (2.1) we may use the inequality

$$\text{largest eigenvalue} = \lambda_1 \geq \mathbf{u}^T \mathbf{C}\, \mathbf{u}, \tag{4}$$

where $\mathbf{u}$ is any vector such that $\mathbf{u}^T\mathbf{u} = \mathbf{1}$. The equality holds if and only if $\mathbf{u}$ is the eigenvector corresponding to λ_1. Therefore, any choice of $\mathbf{u}$ will give a lower bound for λ_1 and thus a lower bound for ρ^{-1}. For regular lattices $\mathbf{u} = \mathbf{n^{-1/2}\,1}$ will typically be a good choice. For the square lattice:

$$\lambda_1^{-1} \leq 0.500 \quad \text{when} \quad \sqrt{n} = 2$$
$$\lambda_1^{-1} \leq 0.259 \quad \text{when} \quad \sqrt{n} = 30$$

suggesting a rather faster rate of convergence to 0.25 than Griffith's results. For smaller lattices, the exact value of λ_1 is readily computed; one should note that the smallest eigvenvalue, λ_n, satisfies

$$\lambda_n \leq \mathbf{u}^T\mathbf{C}\mathbf{u}, \quad \text{for any choice of } \mathbf{u} \text{ with } \quad \mathbf{u}^T\mathbf{u} = \mathbf{1}.$$

However, an effective intuitive choice for $\mathbf{u}$ is more difficult here. For regular lattices, $\lambda_n = -\lambda_1$ works well, as can be seen from equation (1.1); this choice is also made by Griffith. Fortunately whenever $\rho > 0$, as is usually the case, the exact choice for λ_n has little impact on the estimation process.

Approximations to the rest of the Jacobian function are still required, particularly to ensure the accurate assessment of the large sample variance.

Finally, it should be noted that the single parameter case is tractable because of the eigenvalue approach, but that this approach fails for two or more parameters (unless the weighting matrices are orthogonal). However, some progress may be possible using (4). Let the inverse of the covariance matrix be

$$\mathbf{B} = \mathbf{I} - \rho_1\mathbf{C}_1 - \rho_2\mathbf{C}_2.$$

We know that $\mathbf{u^T B u} > \mathbf{0}$ so that approximations to the determinant might be feasible by selecting suitable $\mathbf{u}$ to generate factors like

$$(\mathbf{1} - \rho_1\mathbf{c}_1 - \rho_2\mathbf{c}_2).$$

We know that the determinant is the product of $\mathbf{n}$ such factors. For example, using

$$\mathbf{u} = m^{-1}\mathbf{1}$$

on a regular square lattice with $\mathbf{n} = \mathbf{m}^2$, where $\mathbf{C}_1$ denotes East–West links and $\mathbf{C}_2$ denotes North–South links, produces the factor

$$\mathbf{1} - [\mathbf{2(m-1)/m}]\rho_1 - [\mathbf{2(m-1)/m}]\rho_2,$$

implying, for large $\mathbf{m}$,

$$\rho_1 + \rho_2 \leq 1/2.$$

Whether such an approach produces sensible answers remains to be seen, but Griffith's paper has opened the door to new lines of attack.

References

Anderson, O. (1981) Serial dependence properties of linear processes. *Journal of the Operations Research Society*, **31**, 905-917.

Bartlett, M. (1975) *The Statistical Analysis of Spatial Pattern*. London: Chapman and Hall.

Ripley, B. (1988) *Inference for Spatial Processes*. Cambridge: University Press.

J. Keith Ord, The Pennsylvania State University