Available online at www.sciencedirect.com

ScienceDirect

Procedia Environmental Sciences 26 (2015) 120 – 123

Procedia Environmental Sciences

Spatial Statistics 20ti15: Emerging Patterns

Implementing approximations to extreme eigenvalues and eigenvalues of irregular surface partitionings for use in SAR and CAR models

Daniel Griffith[a], Roger Bivand,[bc1] Yongwan Chun[a]

[a]*University of Texas at Dallas, Richardson, Texas 75080, USA*
[b]*Norwegian School of Economics, 5045 Bergen, Norway;* [c]*Adam Mickiewicz University, 61-680 Poznan, Poland*

Abstract

Good approximations of eigenvalues exist for the regular square and hexagonal tessellations. To complement this situation, spatial scientists need good approximations of eigenvalues for irregular tessellations. Starting from known or approximated extreme eigenvalues, the remaining eigenvalues may be in turn approximated. One reason spatial scientists are interested in eigenvalues is because they are needed to calculate the Jacobian term in the autonormal probability model. If eigenvalues are not needed for model fitting, good approximations are needed to give the interval within which the spatial parameter will lie.

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns committee

Keywords: Extreme eigenvalues; Spatial regression; Rayleigh quotient

1. Introduction

Spatial regression models are fitted in a wide range of disciplines, from political and regional science to epidemiology and ecology; simultaneous (SAR) and conditional (CAR) autoregressive models are among these. Testing for spatial autocorrelation, and the study of methods for specifying and fitting spatial regression models have been central topics in spatial analysis and quantitative geography for many decades, but have not lost their relevance and research interest. Problems can however arise when data sets become large because it is necessary to compute the log determinant of an n by n matrix when optimizing the log-likelihood function, where n is the number of observations. This matrix is $|\mathbf{I} - \rho\mathbf{W}|$, where $|*|$ denotes the determinant of matrix $*$, $\mathbf{I}$ is the identity matrix, ρ is a spatial coefficient, and $\mathbf{W}$ is an n by n matrix of fixed positive finite spatial weights.

* Corresponding author. Tel.: +47-55959355*E-mail address:* roger.bivand@nhh.no

Peer-review under responsibility of Spatial Statistics 2015: Emerging Patterns Committee
doi:10.1016/j.proenv.2015.05.013

For moderate n, the log determinant may be found using the eigenvalues of **W** numerically (Ord) [1]:

$$\ln(|\mathbf{I} - \rho\mathbf{W}|) = \sum \ln(1 - \rho\lambda_i)$$

where λ_i are the eigenvalues of **W**. If the matrix is symmetric, the eigenvalues will have imaginary parts equal to zero, if asymmetric, some imaginary parts may be non-zero (The eigenvalues for an asymmetric matrix that has a symmetric similarity matrix counterpart are all real). For larger n, the log determinant may be found using LU or Cholesky decomposition, or by approximations; see Bivand et al. [2] for a recent review. For large regular grids, the eigenvalues of **W** are known analytically for contiguity defined as non-zero shared boundary length (Ord; Griffith and Sone) [1, 3].

The work on which we report here is part of an on-going initiative to approximate eigenvalues for irregular surface partitionings. Because approximations need to be rescaled using the extreme eigenvalues, attention has first been given to their calculation. This has the beneficial side effect of providing a review of methods of calculating such extreme eigenvalues for **W** configured in different ways. First we will report on methods for calculating extreme eigenvalues and then show how these have been implemented in R with examples and timings, before providing conclusions on this stage of our work.

2. Extreme eigenvalues

If n is small, the eigenproblem may be solved for the spatial weights matrix, removing the need for separate calculation of the extreme eigenvalues. Griffith [4] shows how extreme eigenvalues of a binary symmetric contiguity matrix **C** expressed as an undirected irreducible planar graph may be calculated. The largest eigenvalue is found using the Rayleigh quotient approach.

$$\lambda_1 = lim\,(1'C^{k+1}1)/(1'C^k 1)$$

Although the discussion is framed in terms of an undirected irredicible planar graph, it turns out that this approach to finding the largest eigenvalue is quite robust. See also discussion of use of Rayleigh quotient approaches in Nash [5,6]. Griffith [4] further describes an approach to finding the smallest eigenvalue of **C**. This is extended to the case of the smallest eigenvalue of **W**, the row-standardised spatial weights matrix based on **C**. Since the largest eigenvalue is known to be one by defintion, the lower bound remains to be approximated, except where a result given by Smirnov and Anselin [7] can be applied.

If the data include observations with no neighbours, the graph is no longer irredicible, but the corresponding eigenvalues are zero, and may be ignored. If any subgraphs have two members, the corresponding matrix is cyclical in Smirnov and Anselin's terminology, and the smallest eigenvalue must be -1 by definition. If larger subgraphs, or the graph as a whole, are cyclical --- for every location i, no pair of its neighbours j, k are connected --- the same result holds. A classic case is that rook contiguity on a regular grid meets this condition. Obviously, when the weights matrix is block diagonal, the subsets result in a collection of smaller problems that may be solved one-by-one.

Lehoucq et al. [8] also provide methods implemented in ARPACK for finding extreme eigenvalues of large sparse matrices; it may however be the case that fine-tuning of the number of iterations is required.

In discussion by Kelejian and Prucha [9] and developed in the Stata implementation presented by Drukker et al. [10], the largest eigenvalue of a general binary weights matrix is used to normalise the matrix, in a way reminiscent of treatments in spatial statistics. They suggest the use of a minmax approach, taking the minimum of maxima of row sums and column sums of the weights matrix as an approximation of the largest eigenvalue, but other approximations, such as that given by Griffith [4] seem superior.

3. Implementations

The methods proposed in Griffith [4] have been implemented in the R package spdep based on Fortran code used in the original paper, and with some steps coded in C. We are now in the process of testing edge cases, such as the robustness of these methods when planarity cannot be established directly. We observe that ARPACK code in the rARPACK package is more robust than that in igraph, but that graph analysis tools in this package are useful in finding possible short-cuts when the weights matrix is row-standardised and there are multiple subgraphs (block diagonal weights matrix), where some of these blocks may be cyclical in the Smirnov-Anselin sense. The poster presented at the Spatial Statistics conference will demonstrate using example code, which will be available for download from http://spatial.nhh.no/misc/ss15/extremeEigenvalues.

Figure 1 shows the values taken by the log determinant between its bounds for two spatial weights matrices based on Queen contiguities for 1970 census tracts in Boston, used in the classic study of the influence of air pollution on house prices. It is given here as an indication only.

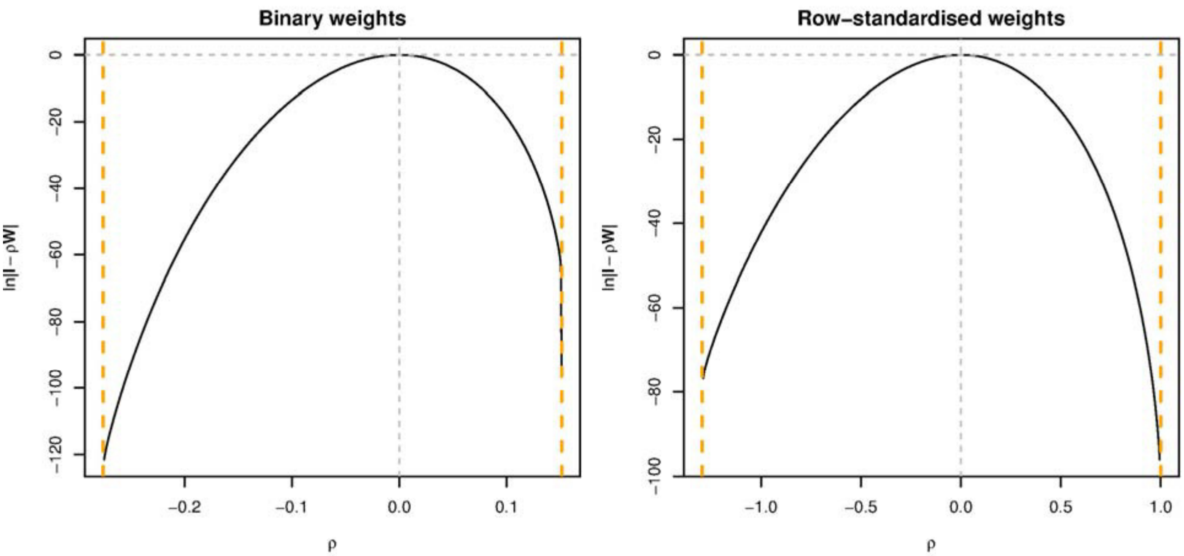

Figure 1: Left panel shows values of the log determinant for values of the spatial coefficient within its bounds (vertical dashed orange lines) for binary weights; right panel shows values of the log determinant for values of the spatial coefficient within its bounds (vertical dashed orange lines) for row-standardised weights, 506 Boston census tracts.

References

1. J. Ord, Estimation methods for models of spatial interaction, Journal of the American Statistical Association 70 (349) (1975) 120-126.
2. R. S. Bivand, J. Hauke, T. Kossowski, Computing the Jacobian in Gaussian spatial autoregressive models: An illustrated comparison of available methods, Geographical Analysis 45 (2) (2013) 150-179.
3. D. A. Griffith, A. Sone, Trade-offs associated with normalizing constant computational simplifications for estimating spatial statistical models, Journal of Statistical Computation and Simulation 51 (2-4) (1995) 165-183.
4. D. A. Griffith, Extreme eigenfunctions of adjacency matrices for planar graphs employed in spatial analyses, Linear Algebra and its Applications 388 (2004) 201-219.
5. J. C. Nash, Compact Numerical Methods for Computers, Adam Hilger, Bristol, 1979.
6. J. C. Nash, Nonlinear Parameter Optimization using R Tools, Wiley, Chichester, 2014.

7. O. Smirnov, L. Anselin, An O(N) parallel method of computing the Log-Jacobian of the variable transformation for models with spatial interaction on a lattice, Computational Statistics & Data Analysis 53 (8) (2009) 2980-2988.
8. R. Lehoucq, D. Sorensen, C. Yang, ARPACK Users' Guide, Society for Industrial and Applied Mathematics, Philadelphia, 1998.
9. H. H. Kelejian, I. R. Prucha, Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances, Journal of Econometrics 157 (1) (2010) 53-67.
10. D. M. Drukker, I. R. P. H Peng, R. Raciborski, Creating and managing spatial-weighting matrices with the spmat command, Stata Journal 13 (2) (2013) 242-286.