BAYESIAN MODEL FOR SPATIAL DEPENDANCE AND PREDICTION OF TUBERCULOSIS

Article · November 2013

1 author:

Perumal Venkatesan
Sri Ramachandra University
222 PUBLICATIONS   **3,463** CITATIONS

Some of the authors of this publication are also working on these related projects:

Machine Learning Approaches for Biological Data mining View project

Survival View project

International Journal of Research In Medical and Health Sciences

BAYESIAN MODEL FOR SPATIAL DEPENDANCE AND PREDICTION OF TUBERCULOSIS

Srinivasan R and Venkatesan P

Dept. of Statistics, National Institute for Research Tuberculosis, (Indian Council of Medical Research), srinivast_r@yahoo.com and venkaticmr@gmail.com

ABSTRACT

In the Disease analysis, there is an interest to find the spatial pattern of disease in specific regions and whether this pattern has any spatial dependence. Kriging is one method used to find the spatial dependence and extrapolate the location of cases from unmeasured locations. But Bayesian Kriging would be more appropriate in the case of tuberculosis risk where we know that other factors are strong predictors in tuberculosis disease. The aim is to study the spatial pattern and spatial dependence of tuberculosis within a Chennai ward population to gain insight into the disease spread and also, to extrapolate the disease location from the unmeasured disease locations. SAS and WinBUGS software were used for spatial analysis of tuberculosis spread. Data was obtained from National Institute for Research in Tuberculosis for Chennai district. The result reveals that Kriging has significantly improved the prediction of tuberculosis risk in parts of the Chennai city. The GIS system proves to be a friendly interface for spatial and a spatial information retrieval, which supports users with all type of statistical analysis GIS model.

Keywords: Bayesian Kriging, Autocorrelation, Moran's I, Geary's C.

Introduction

Kriging is a technique used in the analysis of spatial data. The data from the measured location can be used to estimate the variable at the location where it had not been measured. This extrapolation from measured location to unmeasured location is called kriging. Measurements of variable at a set of points in a region are used to extrapolate points in the region where the variable was not measured outside the region that we believe will behave similarly. In both cases, we will need to first fit a variogram model to our data. The three major functions used in spatial statistics for describing the spatial correlation of observations are the correlogram, the covariance, and the semi-variogram. The last is also more simply called the variogram. The variogram is the key function in spatial statistics as it is used to fit a model of the spatial correlation of the data.

Observations made at different locations may not be independent. For example, measurements made at nearby locations may be closer in value than measurements made at locations farther apart (Tobler law). This phenomenon is called spatial autocorrelation which measures the correlation of a variable with itself through space. This spatial autocorrelation value can be positive or negative. Positive spatial autocorrelation occurs when similar values occur near one another. Negative spatial autocorrelation occurs when dissimilar values occur near one another. The two classic works reviewing and extending spatial statistical theory are given by Cliff and Ord (1981), whom have motivated research involving spatial autoregression, and Cressie (1993) has summarized research involving geostatistics. Geostatistics uses variance-covariance matrix while spatial autocorrelation uses inverse of this matrix. Griffith and Layne (1999) among others, shows links between geostatistics and spatial auto regression.

Moran's I is one method used to find the autocorrelation of disease based on the location which is one of the oldest indicators of spatial autocorrelation (Moran, 1950). It measures the strength of spatial autocorrelation in a map. For calculating spatial autocorrelation, neighboring values can be identified by an n x n binary geographic weights matrix, say, C; if two locations are neighbors, then $c_{ij} = 1$, and if not, then $c_{ij} = 0$, in which two areal units are deemed neighbors if they

share a common non-zero length boundary. Test of significance can be used for testing independence. There are many ways to approach the analysis of the spatial pattern of tuberculosis and HIV. If there is any systematic pattern in the spatial distribution, it is said to be spatially autocorrelated. One approach is to define disease in a ward to be close to one another, and then determine, whether pattern may have similar characteristics. Once the spatial correlation structure of a variable has been identified, the data from the measured locations can be used to estimate the spatial dependence based on location is significant or not (Banerjee et al., 2004).

Bayesian approach brings additional flexibility to the classical prediction framework outlined above. First of all, the issue of incorporating uncertainty in covariance parameters follows directly from posterior inference. More specifically, Bayesian prediction derives from the posterior predictive distribution which integrates over the posterior distribution of all model parameters, In the Bayesian context, transformations are less problematic, as predictions derive from the posterior predictive distribution rather than a necessarily linear combination of observations. This is particularly evident in MCMC implementation where a sample from the posterior distribution of model parameters may be transformed to a sample from the posterior distribution of the transformed parameters with little effort.

Material and methods:

The datasets are taken from National Institute for Research in Tuberculosis (NIRT), formerly TRC, which is conducting clinical trials for Tuberculosis and HIV in Chennai and its suburbs since 1956. The patients registered during 2004 to 2006 for an ongoing trial were considered for this study. Chennai had 155 wards and for each ward the total number of TB cases recorded between 2004 and 2006 were identified. For variogram analysis, 28 wards of Chennai district were selected for our study for which the locations of 72 cases were geographically marked through their co-ordinates in the Chennai map.

Kriging analysis was carried out using SAS software by dividing the whole area into some 100 by 100 grid matrix and prediction is calculated using ordinary kriging method. For Bayesian kriging, WinBUGS software was used for prediction of certain locations based on available information about other location. The spatial prediction permits spatial interpolation and prediction in WinBUGS. The data for this work consist of values of SMR, and coordinates of x and y of the each wards. The spatial.exp function allows the fitting of a fully parameterized

covariance function within a multivariate normal distributional model. Spatial.unipred provides a method of predicting values of the fitted surface at unsampled locations.

Results:

The diagonal covariance of the variogram follows the very general increasing then flattening shape of our data. Also, the variogram increases with distance at small distances and then levels off after certain point. This general shape is suggestive of a spatial correlation that is positive and strong at small distances and becomes less so as distances increase until reaching a certain distance

" Figure 1 about here"

In Fig.1 the graph of three theoretical variograms (Gaussian, exponential and spherical) along with the observed one are presented. From the Fig. we observe that all three theoretical variograms follow the very general increasing then flattening shape. The Gaussian variogram appears to most closely match the data. All three variograms increase with distance at small distances and then levels off. This graph also gives us a chance to see how these three theoretical variograms differ in shape: exponential increases gradually and is concave over the range; spherical features a sharp increase and a quick leveling off; Gaussian offers a compromise between the two. Spatial dependence of these was modeled using the constructed variogram. All the variogram shows spatial dependence exist upto 40 meter in Chennai city. Table 1 gives the variogram fit statistics for the three models.

"Table 1 about here"

From the table 1, we infer that the spherical model has lower deviance, AIC, BIC when compared with to exponential and Gaussian models. Hence, we have chosen the spherical model for further analysis. The variogram shows that there is some evidence of spatial correlation over short distance of below 40 m.

Kriging

Table 2 shows the observed and predicted values of the SMR in Chennai wards. The Bayesian Kriging prediction gave results close to the observed SMR. Also the Bayesian approaches resulted in lower SE

"Table 2 about here"

Spatial Autocorrelation

International Journal of Research In Medical and Health Sciences

In our study, nearby areas are more alike, and it indicates positive spatial autocorrelation and there is no negative autocorrelation or Random patterns exhibit in this area. The results are presented in Fig. 2.

"Figure 2 about here"

The Moran's spatial autocorrelation of Chennai city is MC: 0.32 and statistically significant also. The result reveals that the high risk area surrounded by high risk and moderately high risk area and low risk area surrounded by low risk and risk free area and some wards show no autocorrelation between disease and spatial pattern. It is reflected in map also.

Conclusion

The GIS system proves to be a friendly interface for spatial and a spatial information retrieval, which supports users with all type of statistical analysis GIS model. Spatial dependence exists between small distances of tuberculosis cases found in Chennai wards. The Deviance of Spherical model is less and closely matching our data set next to exponential and Gaussian model. This is consistent with the findings of the variogram graph results. The variogram reveals that there is a correlation exists over the space.

Kriging has significantly improved the prediction of tuberculosis risk in parts of the Chennai city, however, given that the data used for obtaining the model are not a random sample of the population or a spatially well distributed set of sampling points, and extrapolating the predicted risk to points outside the data set closely matched with our observed SMR. A concern with spatial data is the potential for spatial correlation in the observations, which could lead to incorrect estimates. An infectious disease that is heavily associated with other variables is likely to be spatially clustered. The model derived here explains some of the spatial dependence of tuberculosis risk, with significant spatial correlation, particularly over short distances of under 40 m and is confirmed by the variogram method.

Bayesian Kriging would be more appropriate in the case of tuberculosis risk where we know that other factors are strong predictors. The Moran's spatial autocorrelation of Chennai city is 32% and statistically significant also. The result reveals that the high risk area surrounded by high and moderately high risk areas and low risk areas surrounded by low risk areas and some wards shows no autocorrelation between disease pattern.

REFERENCES

1. Banerjee S, Carlin B and Gelfand A E (2004): Hierarchical Modeling and Analysis for Spatial Data. Boca Raton: Chapman & Hall.

2. Carrat F and Valleron A J (1992): Epidemiologic mapping using the kriging method: application to an influenza-like illness epidemic in France, American Journal of Epidemiology, 135: 1293-1300.

3. Cliff A and Ord J (1981): Spatial Processes. Pion, London.

4. Cressie N (1985): Fitting variogram models by weighted least squares, Mathematical Geology, 5: 563-586.

5. Cressie N (1986): Kriging Non-stationary Data. Journal of the American Statistical Association, 81: 625–634.

6. Cressie N (1993): Statistics for spatial data. New York: John Wiley & Sons.

7. Diggle P (1993): Statistical Analysis of Spatial Point Patterns. New York: Academic Press.

8. Diggle P, Morris S and Wakefield J (2000): Point-source modelling using matched case-control data. Biostatistics, 1: 1-17.

9. Diggle P, Tawn J A and Moyeed R A (1998): Model-based geostatistics. Journal of Royal Statistical Society, Series-C (Applied Statistics), 47: 299

10. Gelfand A Diggle P and Guttorp P (2010): Handbook of spatial statistics, Chapman & Hall / CRC.

11. Isaaks E and Srivastava R (1989): An Introduction to Applied Geostatistics. Oxford: Oxford University Press.

12. Lawson A (1993): On the analysis of mortality events around a prespecified fixed point. Journal of the Royal Statistical Society, Series-A, 156: 363-377.

13. Matheron G (1963): Principles of geostatistics, Economic Geology, 58: 1246-1266.

14. Pickle L, Mungiole M, Jones G and White A (1999): Exploring spatial patterns of mortality: the new atlas of United States mortality. Statistics in Medicine 18: 3211- 3220.

15. Ripley B (1981): Spatial Statistics. New York: Wiley.

16. Shapiro A and Botha J (1991): Variogram fitting with a general class of conditionally nonnegative definite functions. Computational Statistics and Data Analysis, 11: 87-96.

17. Stein A and Corsten L (1991): Universal kriging and cokriging as a regression procedure. Biometrics, 47: 575-587.

18. Stein A, Van Eijnbergen, A C and Barendregt, L G (1991): Cokriging nonstationary data. Mathematical Geology, 23: 703-719.

19. Stein M (1999a): Interpolation of Spatial Data: Some Theory for Kriging. New York: Springer-Verlag.

20. Venkatesan P and Srinivasan R (2010): Modeling the variogram of tuberculosis in Chennai ward, Indian Journal of Science and Technology. 3 : 167-69.

21. Wakefield J and Morris S (2001): The Bayesian modeling of disease risk in relation to a point source. Journal of American Statistical Association, 96: 77-91.
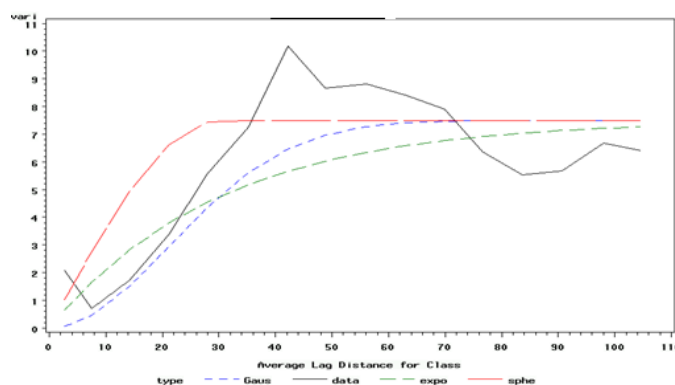
Figure 1 Variogram: Observed and theoretical variogram for Chennai wards

Table 1 Fit Statistics of Variogram

Statistic	Diagonal	Spherical	Exponential	Gaussian
Deviance	298.1	199.9	204.2	255.2
AIC	300.1	203.9	208.2	259.2
BIC	302.2	208.2	212.5	263.4

Table 2 Kriging (prediction) Estimates

Wards	SMR	Ordinary Kriging		Bayesian Kriging	
		Prediction	Std Error	Prediction	Std Error
Ward1	0.469	0.359	0.11	0.431	0.038
Ward 2	1.611	1.181	0.43	1.591	0.02
Ward 3	0.628	0.428	0.2	0.614	0.014
Ward 4	0.646	0.87	0.224	0.635	0.011
Ward 5	0.352	0.232	0.12	0.362	0.01
Ward 6	0.609	0.94	0.331	0.611	0.002
Ward 7	2.492	0.942	1.55	2.123	0.369
Ward 8	0.858	0.488	0.37	0.812	0.046
Ward 9	0.701	0.211	0.49	0.712	0.011

5

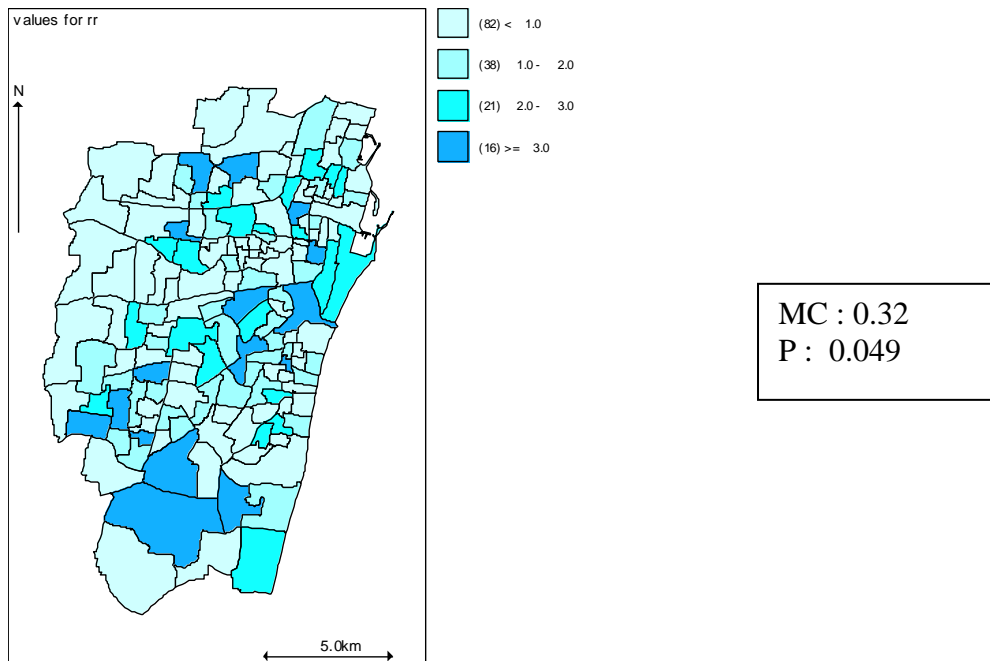| Ward 10	1.01	1.31	0.3	1.11	0.1

MC : 0.32
P : 0.049

Figure 2 Measure of Spatial Autocorrelation for Chennai wards

6