

A Statistical Model to Evaluate the Relationship between Disease Risk and Environmental Point Source: Some Empirical Results.

Alcuni Risultati Empirici per Valutare l'Associazione tra Rischio di Malattia e Fonti di Inquinamento Atmosferiche.

Eugenia Nissi - Annalina Sarra

Dipartimento di Metodi Quantitativi e Teoria Economica

Viale Pindaro, 42 -65127 Pescara

nessi@dmqte.unich.it - asarra@dmqte.unich.it

Riassunto: Negli ultimi anni, si è osservato un interesse crescente verso i metodi statistici che consentono di valutare l'associazione fra rischio di malattia e fonti di inquinamento atmosferico. L'obiettivo del presente lavoro è quello di analizzare le caratteristiche di un modello ad effetti direzionali che consenta di descrivere il rischio di malattia in prossimità di punti sorgenti. Il modello rappresenta un'estensione di modelli proposti da Diggle per studi di tipo caso-controllo in una prefissata area di studio. Attraverso un esperimento simulativo di tipo Monte Carlo viene esaminata l'adeguatezza di utilizzare l'inversa della matrice di informazione osservata di Fisher come approssimazione della matrice di varianza dei parametri del modello proposto.

Keywords: spatial point process, binary regression, environmental epidemiology, directional plume model, odds ratio.

1.Introduction

As known, environmental epidemiology regards the study of disease patterns in relation to environmental factors. In recent years there has been growing interest in the development of statistical methods useful in detection of patterns of health events associated with pollution sources. Many epidemiological studies have traditionally used simple, proxy measure of exposure, such as the distance from point source. Our objective in this paper is to introduce a parametric modelling approach to the study of disease risk near point source.

The paper is organized as follow: in section 2 we describe an extension of the parametric modelling framework developed by Diggle and Rowlingson (1994), for case-control data, to investigate the raised risk around putative source of environmental pollution. As this model ignores directional effects, we propose a more flexible anisotropic model that allows for differential risk at different orientations: "the directional plume model".

In section 3 we perform a simulation study to investigate if the usual asymptotic approximations for likelihood-based inference are reliable.

2. The directional plume model.

Our starting point is an inhomogeneous Poisson point process model for the spatial variation in the incidence of cases and controls in a prefixed geographical region, as proposed by Diggle (1990). In this study we refer to case-control data: exact locations of individuals with and without disease are available. The set of controls are independent random samples from the same underlying population at risk. Owing to the presence of the point source, we have to relax the stationary assumption. Then, it can usually be assumed that the cases are governed by a Heterogeneous Poisson Process, with intensity function given by

$$\lambda_1(x) = \rho \lambda_0(x) f\{x_0; \theta\} \quad (1)$$

where $\lambda_0(x)$ is the controls intensity function which reflects the population structure of the area; ρ is a scaling parameter which reflects the overall prevalence of disease relative to the number of controls; $f\{x_0; \theta\}$ describes the relationship between disease and proximity to the point source and x_0 is the location of the point source. A more detailed description of the method can be found in Diggle (1990), Diggle & Rowlingson (1994). However it is dangerous to pursue only based distance analysis when considerable directional effects are present. For instance, the wind regime can easily produce considerable difference in exposure in different directions. A more general class of model to parameterise the association between disease risk and point source has been introduced by Lawson (1993). Such a class allows for non-monotonic distance risk relationship and also for directional effects. Lawson (1993) proposes a parametric approach in which directional effects are modelled through log-linear sine and cosine terms. The directional plume model is an adaptation of the point process method suggested by Diggle & Rowlingson (1994) in which the intensity function depends on radial and directional components. By using polar coordinates (d, θ) , we can assume that the case intensity function has a multiplicative decomposition into

$$\lambda_1(x) = \rho \lambda_0(x) \left(1 + \alpha \exp \left(- \left[d \exp \{k \cos(\theta - \phi)\} / \beta \right]^2 \right) \right) \quad (2)$$

where $d = \|x - x_0\|$ ($\|\cdot\|$ is the distance) and θ is the angle between the event location and source location.

Instead the model parameters are :

α which represents the elevation in risk at source; β which indicates the rate of decay of risk with distance source; ϕ is the direction of plume and k is the degree of directional concentration of plume.

The proposed model has the advantage to be a parsimonious representation of directional effects when an outfall around an air pollution source exhibits a plume. In Nissi & Sarra (2003) can be found further details about the maximum likelihood estimation of model parameters.

3. Simulation study: some results.

The focus of interest for assessing association between events and the source is inference regarding parameters in $f\{x_0; \boldsymbol{\theta}\}$, treating ρ as a nuisance parameter. As previous noted, the model proposed, which allows for directional effects, is an augmented version of the conditional logistic model of Diggle & Rowlingson (1994). They derived an exact conditional likelihood function assuming that conditionally on the locations of all cases and controls, the labels of individuals, divided into cases and controls, constitute a set of mutually independent Bernoulli trials with the probability that an individual at x is a case $p(x)$ is given by

$$\frac{\lambda_1(x)}{\lambda_0(x) + \lambda_1(x)} = \frac{\rho f(x_0; \boldsymbol{\theta})}{1 + \rho f(x_0; \boldsymbol{\theta})}$$

Following their approach, $\lambda_0(x)$, the background risk, is “conditioned out” of the model and need not be estimated. The vector parameters $\boldsymbol{\theta}$ can be estimated by maximum likelihood estimation of odds ratio $\frac{p(x)}{1 - p(x)}$.

So the log-likelihood function for ρ and $\boldsymbol{\theta}$ can be expressed as

$$L(\rho; \boldsymbol{\theta}) = n \log \rho + \sum_{i=1}^n \log f(x_0, \boldsymbol{\theta}) - \sum_{i=1}^{n+m} \log \{1 + \rho f(x_0, \boldsymbol{\theta})\} \quad (3)$$

where n =numbers of cases; m =numbers of controls and $f(x_0, \boldsymbol{\theta})$ is the directional plume model, specified as above. The parameter estimates can be obtained by numerical maximization of equation (3). Their standard errors can be obtained by evaluation of the observed information matrix.

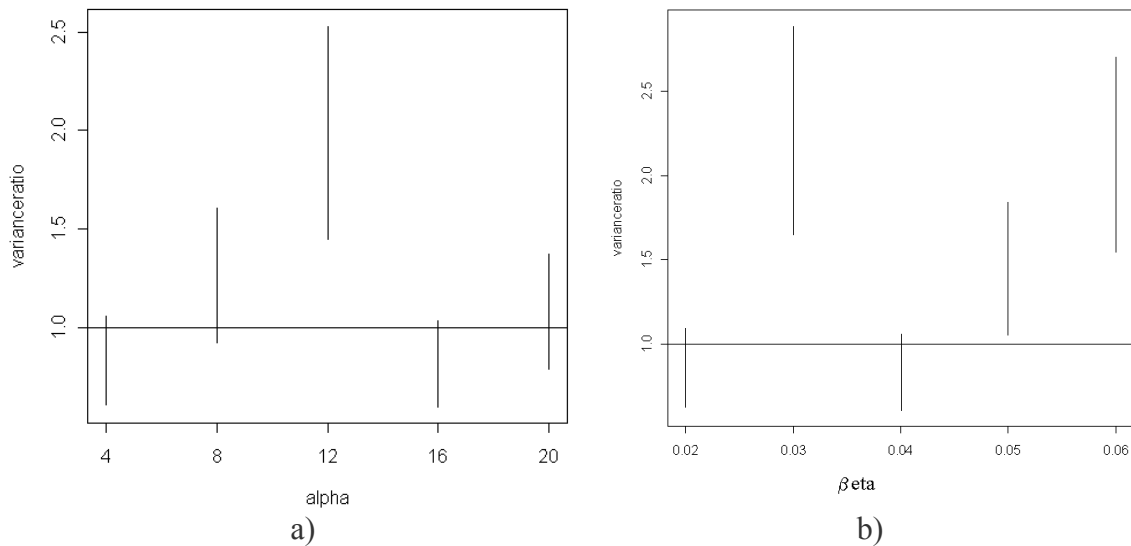
Different choices about the parametric form of the function $f(x_0, \boldsymbol{\theta})$, may or may not assure that the usual asymptotic approximations for likelihood-based inference are reliable.

Hence in this paper we perform a simulation study to check if the nominal variances derived from numerical estimates of the Hessian matrix can be conveniently used. The simulation scheme involves the definition of the study region A as a unit radius disc centred at the origin, which is also the source location. In the simulation procedure we assumed that the n locations of disease cases are realizations of inhomogeneous poisson process with intensity function given by expression (2). We also fixed $\rho = 0.1$, $\lambda_0(x) = \lambda_0$, a constant, and $n+m=1000$. To generate the inhomogeneous Poisson process we refer to the sampling rejection algorithm (Lewis and Shelder, 1979). Our goal is the comparison between the nominal variances of the maximum likelihood estimators (calculated by the Hessian matrix) and the empirical value, computed from an application of the maximum likelihood method to replicate simulations of the same model. The results are obtained by varying one parameter and fixing the others. For each combination of the model parameters (α, β, ϕ, k) , we simulate 100 independent realizations of the fitted model. The 100 simulations generate 100 values $\hat{\theta}_{ij}$

$i=1,\dots,100$, $j=1,\dots,4$ and associated approximate variance v_{ij} (obtained by evaluation of the observed information matrix). From these, we calculate s_j^2 , the sample variance of $\hat{\theta}_{ij}$ and $\bar{v}_{ij}$ the sample mean of v_{ij} . The sampling distribution of s_j^2 is proportional to χ^2 on 99 degrees of freedom (Diggle & Rowlingson, 1994). The confidence limits for the variance ratio are 0.975 and 0.025 quantiles of χ^2_{99} .

In our simulation study we have found that the comparison between the nominal variance (obtained by the Hessian matrix) and that one computed from an application of the maximum likelihood method to replicate simulations of the same model is unreliable. For the model parameters (α and β) the observed information variance seems to be a good approximation to the nominal variance. For the remaining parameters the variance ratio is far to be 1. These results recommend the use of Monte Carlo estimates of standard errors.

Figure 3.1. Variance ratio for varying α and fixed $\beta=0.05$, $\phi=4.18$ and $k=3$ (a) Variance ratio for varying β and fixed $\alpha=4$, $\phi=4.18$ and $k=3$



References

- Diggle P.J. (1990) A point process Modelling Approach to raised Incidence of a Rare Phenomenon in the vicinity of a pre-specified point, *Journal of the Royal Statistical Society. Series A* 153, pp. 349-362.
- Diggle P.J, Rowlingson, B.S. (1994) A conditional approach to point process modelling of elevated risk, *Journal of the Royal Statistical Society. Series A* 157, pp. 433-440.
- Lawson A.B (1993) On the analysis of mortality events around a prespecified fixed point, *Journal of the Royal Statistical Society. Series A* 156, pp. 363-377.
- Lewis P.A.W , Shelder, G.S. (1979).Simulation of non-homogenous Poisson processes by thinning, *Naval Research Logistic Quarterly*, 26, pp. 403-413.
- Nissi E., Sarra, A. (2003) Un modello con effetti direzionali per l'analisi del rischio di malattia in prossimità di punti sorgenti, *Atti del V° Congresso Nazionale della SIB*.