Empirical and Full Bayes estimators for disease mapping

V. Gómez-Rubio[1] and A. López-Quílez[2]

[1] Epidemiology and Public Health, Imperial College London, London, United Kingdom
[2] Estadística e Invesitación Operativa, Universidad de Valencia, Valencia, Spain

1 Introduction

Disease mapping tackles the problem of providing a description of the geographical variation of disease by means of estimating disease risk in small areas. Given the number of factors that may effect the spatial variation of the risk, it is very important to take into account possible risk factors and spatial interaction between areas. If data have been collected at different times, temporal effects can be explored too.
Hence, the estimation of the risk of disease can lead to the formulation of complex spatial or spatio-temporal models whose parameters need to be estimated. Hierarchical Bayesian models can handle different types of effects and sources of variation, and provide a unified way of dealing with the uncertainty of the estimation of the parameters of the model. In a Full Bayes (FB) setting, all the parameters are assigned prior distribution whose parameters are also assigned hiperprior distributions to cope with their possible variability. The estimation procedures can be done by using Markov Chain Monte Carlo simulation techniques. Alternatively, the parameters of the model can be estimated by maximising their posterior distribution. This is known as Empirical Bayes (EB) estimation.

2 Empirical Bayes Estimators

Clayton and Kaldor (1987) proposed two simple Empirical Bayes

(EB) estimators based on the Poisson-Gamma and log-Normal model that borrowed information from all the small areas but without considering the spatial configuration of the small areas. Marshall (1991) proposed an EB estimator without assuming a prior distribution for the relative risks but only their priori mean and variance. All these EB estimators produce estimates of the disease risk that are a compromise between the direct estimate of the risk (usually, the Standardised Mortality Ratio) and the prior mean of all the relative risks.
Marshall (1991) also developed another EB that used only local information and, hence, produce estimates that are shrunk towards the local mean.
EB estimators have been criticised because they provide estimates that they are too shrunk towards the global (or local) mean. Furthermore, this is also reflected in the distribution of the set of estimates which has less variability than that of the direct estimates. Devine and Louis (1994) considered *constrained* EB estimators that produce estimates whose joint variation is similar to that of the raw data.

3 Full Bayes analysis for disease mapping

Best et al. (2005) summarise the most recent Hierarchical Bayesian models that are used for disease mapping using Full Bayes (FB) estimation. Knorr-Held (2000) describes four possible interactions between space and time that can be used to construct spatio-temporal models for disease risk. Inference is made by using Markov Chain Monte Carlo techniques that provide an estimate of the posterior distribution of the parameters of the model.
The spatial interaction is usually modelled by mean of a Conditional Autoregressive model(Waller and Gotway, 2004), so that only the effects of nearby areas are included. In a similar way, the temporal effects use an Autoregressive model that measures the relation of the values in one area at different times. Finally, both space and time effects can interact. Unfortunately, considering some or all these effects at the same time may cause problems

of identifiability of the effects, and further assumptions may be needed.

4 Empirical versus Full Bayes estimation

Bernardinelli and Montomoli (1992) made a comparison of EB and FB estimation procedures with models that dealt with the geographical variation of disease. As they point out, the FB is preferable because it considers the uncertainty of the parameters of the model whereas EB estimation conditions the estimation on point estimates of the parameters of the model. Hence, EB estimates may be less accurate. On the other hand, FB procedures usually are more computer intensive.
Nevertheless, EB estimators have probed useful, for example, to provide initial assessment of disease risk in the context of Geographical Information Systems for Spatial Epidemiology, where large amounts of data are managed and the algorithms used must be rapid and not very computer intensive (Aylin et al., 1999). If the exploratory analysis indicates that there is a strong risk variation between areas, a FB estimation can be used to detect clusters of disease, for example.
The aim of this work is to compare different EB estimators and FB analysis for spatio-temporal-models. We describe how time can be incorporated following the guidelines proposed by Knorr-Held (2000) to produce spatio-temporal EB estimators.

Bibliography

Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G. and Elliot, P. (1999). A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The U.K. Small Area Health Statistics Unit. *Journal of Public Health Medicine*, **21**(3):289-298.

Bernardinelli, L. and Montomoli, C. Empirical bayes versus fully bayesian analysis of geographical variation in disease risk. *Statistics in Medicine*, **11**:983–1007.

Best, N., Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* **14**(1), 35–59.

Clayton, D. and Kaldor, J. Empirical bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, **43**:671–681.

Devine, O.J. and Louis, T.A. A constrained empirical bayes estimator for incidence rates in areas with small populations. *Statistics in Medicine*, **13**:1119–1133.

Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine* **19**, 2555–2567.

Marshall, R.J. (1991). Mapping disease and mortality rates using empirical bayes estimators. *Applied Statistics*, **40**(2):283–294.

Waller, L.A. and Gotway, C.A. (2004). *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, Hoboken, New Jersey.