Bayesian hierarchical models : An analysis of Portugal road accident data [1]

Conceição Ribeiro
ISE-UALG/CEAUL, cribeiro@ualg.pt

Antónia Amaral Turkman
DEIO-FCUL/CEAUL

João Lourenço Cardoso
LNEC

Abstract: In this work Bayesian hierarchical models are applied to road accident data at a county level, in Portugal, from 2000 to 2007. The objective of the study is to build model-based risk maps for road accidents at county level and to perform an analysis of association between road accidents and potential risk factors, through the inclusion of ecological covariates in the model.

Keywords: Bayesian models; Small Area; Road Safety.

1 Introduction

Investigation into risk factors relating to road safety and transport plays an important role in road accident analysis and prevention. Heterogeneity in road accidents can be related to a range of factors, in particular at a small area level. Bayesian hierarchical models allow the incorporation of spatial and temporal effects through prior information and enable ecological analyses of associations between road accidents and potential risk factors over aggregated areas. One objective is to explore and to examine potential associations between road accidents and regional characteristics. Here we only consider three covariates namely the road length, population county size and county area.

2 Materials and Methods

Consider $Y_{ij} \sim Poisson(E_{ij}\theta_{ij})$, where, for each county i and each year j, Y_{ij} is the observerd number of fatal and severe injury crashes, θ_{ij} is the relative risk, E_{ij} is the expected number of fatal and severe injury crashes, for a constant incidence rate

[1] This work is partially sponsored by national funds through FCT - Fundação para a Ciência e a Tecnologia under the project PEst-OE/MAT/UI0006/2011 and by SFRH/PROTEC/49226/2008 PhD grant.

across all 278 counties of Portugal and all 8 years,

$$E_{ij} = N_{ij}\bar{r} = N_{ij}\frac{\sum_i \sum_j Y_{ij}}{\sum_i \sum_j N_{ij}} \tag{1}$$

with N_{ij} the number of vehicles insured in county i, in year j.

Assume a spatio-temporal model for the relative risk, namely

$$\log(\theta_{ij}) = b_0 + bx_{ij} + u_i + v_i + (\gamma + \delta_i)t_j \tag{2}$$

where b_0, bx_{ij} are the fixed effects,with b_0 the intercept, x_{ij} a vector of covariates, b a vector of fixed effect parameters; u_i are random effects acounting for spatial heterogeneity and v_i are random effects accounting for unstructured heterogeneity. We also assume that u_i and v_i are mutually independent with priors $v_i \overset{iid}{\sim} Normal(0, \sigma_v^2)$ and the $u_i \sim CAR$, respectively; γt_j is a linear trend term in time t_j, δ_i is an interaction random effect between space and time, with prior $\delta_i \sim CAR$. We also assumed the following diffuse priors for the hyperparameters: b_0, b, σ_u^2, and σ_v^2 mutually independent with $b_0, b, \gamma \sim Normal(0, 1000)$, $(\sigma_u^2)^{-1}, (\sigma_v^2)^{-1}, (\sigma_\delta^2)^{-1} \sim Gamma(0.5, 0.0005)$.

The models were applied to road accident data in 278 counties of Portugal, from 2000 to 2007 and were implemented using WinBUGS, (Spiegelhalter et al.,1999) and its add-on program GeoBUGS, (Thomas et al., 2004), and using R-INLA, (Rue and Martino, 2009).The covariates used were geographical area-A, in Km^2, population size-P, in number of inhabitants, road lenght-L, in meters, by county and year.

3 Results

Eight spatial-temporal models are implemented. Model 1 without covariates -1-. Models 2 to 4, incorporate one covariate, model 2 includes the road length, 2-(L), model 3 includes the population size, 3-(P), and model 4 includes the area, 4-(A). Models 5 to 7 incorporate two covariates, model 5 includes road length and population size, 5-(L+P), model 6 includes road length and area, 6-(L+A), and model 7 includes population size and area, 7-(P+A). Finally, model 8 incorporates the three covariates, 8-(L+P+A).

Results obtained using INLA and WINBUGS are very similar, with the advantage of INLA taking much less time to run. Model choice is done using DIC (see table 1); accordingly model 2, which has a smaller value for DIC, is chosen to produce maps to display the posterior expected relative risk. Figure 1 display, on the left side, the observed average of fatal and severe injury crashes along the years under study and on the right side the expected relative risks obtained using model 2.

ST Models		1	2-(L)	3-(P)	4-(A)
DIC	INLA	12799.9	12796.1	12800.2	12803.6
	WB	12799.8	12796.0	12799.6	12803.2
ST Models		5-(L+P)	6-(L+A)	7-(P+A)	8-(L+P+A)
DIC	INLA	12796.7	12796.7	12803.8	12797.0
	WB	12796.7	12797.0	12803.6	12797.3

Table 1: DIC values

		mean	sd	0.025q	0.975q
Fixed effects:	b_0	0.03	0.05	0.02	0.04
	b_L	1.6e-06	4.9e-07	6.5e-07	2.6e-06
	γ	-0.088	0.003	-0.095	-0.082
Variance of random effects:	σ_u^2	0.34	0.07	0.18	0.55
	σ_v^2	0.07	0.02	0.04	0.13
	σ_δ^2	0.007	0.001	0.004	0.011

Table 2: Summary statistics for the parameters in model 2

Figure 1: Portugal: Average of fatal and severe injury crashes and posterior expected relative risks for model 2 in 2000

Apparently the model is not able to capture well the risk of accident on the northwest coast line, indicating that other covariates should be considered to be included in the analysis. Summary statistics for the fixed effects and the variance of the random effects for model 2, in table 2, show that road length can be a factor associated with higher risk of accident. The time trend effect being negative may be an indicator that the number of fatal and severe injury crashes decreased over the study period. The analysis of the variance of the random effects shows that the variability of the relative risk is attributed more to spatial-structured effects than to the uncorrelated heterogeneity or to the space-time interaction.

4 Concluding remarks

This is a preliminary analysis, as we are well aware that there are potential risk factors that are not accounted for in this study. These include socioeconomic factors such as age cohorts, sex cohorts, levels of poverty and employment; transportation-related factors such as road type, road curvature, traffic flow, traffic speed, violation of traffic rules, number of vehicle-kilometers traveled, and environmental factors such as total precipitation, number of rainy days per year, land use, size of rural and urban areas.

References

Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., and Songini, M. (1995) Bayesian analysis of space-time variation in disease risk, *Statistics in Medicine*, 14, 2433-2443.

Ghosh, M., Natarajan, K., Waller, L.A., and Kim, D. (1999) Hierarchical bayes GLMs for the analysis of spatial data: An application to disease mapping, *Journal of Statistical Planning and Inference*, 75(2), 305- 318.

Miaou, S.-P., Song, J.J., and Mallick, B.K. (2003) Roadway traffic crash mapping: A space-time modeling approach, *Journal of Transportation and Statistics*, 6(1), 33-57.

Rue, H. and Martino, S. (2009) Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations, *Journal of the Royal Statistical Society B*, 71(2), 319-392.

Spiegelhalter, D.J., Thomas, A., and Best, N.G. (1999) WinBUGS Version 1.2 User Manual, Technical report, MRC Biostatistics Unit.

Thomas, A., Best, N., Lunn, D., Arnold, R., and Spiegelhalter, D. (2004). GeoBUGS User Manual Version 1.2, Technical report, Department of Epidemiology and Public Health of Imperial College at St Mary's Hospital London.