Bym: An example of disease mapping with covariate

Larynx cancer mortality counts are observed in the 544 district of Germany from 1986 to 1990. We assume the data to be conditionally independent Poisson random variables with mean $E_i \exp(\eta_i)$, where E_i is fixed and accounts for demographic variation, and η_i is the log-relative risk. Together with the counts, for each district, the level of smoking consumption c is registered.

The model for η_i takes the following form

$$\eta_i = \mu + f_s(s_i) + f(c_i) + u_i \tag{1}$$

where $f_s(\cdot)$ is the spatial effect and u_i is the unstructured random effect.

The prior model for $\boldsymbol{f}_s = (f(0), \ldots, f(s), \ldots, f(S-1))$ implemented in the `inla` program is a simple (but most often used) intrinsic GMRF model, see [Rue and Held, 2005, Ch. 3], defined as:

$$f_s(s)|f_s(s'), s \neq s', \lambda_s \sim \mathcal{N}(\frac{1}{n_s}\sum_{s\sim s'} f_s(s'), \frac{1}{n_s\lambda_s}) \tag{2}$$

where n_s is the number of neighbours of site s, $s \sim s'$ indicates that the two sites s and s' are neighbours. λ_s is the unknown precision parameter.

The remaining term in (1), $f(c_i)$, is the unknown effect of of the exposure covariate which assumes value c_i for observation i. The effect of covariate c is modelled as a smooth function $f(\cdot)$ parametrised as unknown values $\boldsymbol{f} = (f_0, \ldots, f_{m-1})^T$ at $m = 100$ equidistant values of c_i. We have scaled the covariate values so that they belong to the interval $[0, 10]$. The vector $\boldsymbol{f}$ is modelled with a second-order random walk (RW2) prior with unknown precision λ_f. A sum-to-zero constraint is imposed on $\boldsymbol{f}_s$ and $\boldsymbol{f}$ separate out the spatial effect and the effect of the covariate from the common mean μ.

The model has three hyperparameters $\boldsymbol{\theta} = (\log\lambda_s, \log\lambda_f, \log\lambda_\eta)$. Following [Rue et al., 2007] we assign a vague LogGamma prior to each element of $\boldsymbol{\theta}$.

Linear effect for the covariate

An alternative model is to assume a linear effect for the covariate c:

$$\eta_i = \mu + f_s(s_i) + \beta c_i + u_i \tag{3}$$

In this case the numer of hyperparameters is reduced to tw, namely $\boldsymbol{\theta} = (\log\lambda_s, \log\lambda_\eta)$

References

[Rue and Held, 2005] Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

[Rue et al., 2007] Rue, H., Martino, S., and Chopin, N. (2007). Approximate bayesian inference for latent gaussian models using integrated nested laplace approximations. Statistics Report No. 1, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim, Norway.