Supplementary material to "Measuring the vulnerability of the Uruguayan population to vector-borne diseases via spatially hierarchical factor models" by Lopes, Schmidt, Salazar, Goméz and Achkar

Supplement A: MCMC and model selection

Contents

1 MCMC schemes **2**

1.1 Spatially hierarchical factor model 2

1.2 Unstructured hierarchical factor model 3

2 Model Selection **4**

2.1 Expected posterior deviation 4

2.2 Deviance information criterion 4

2.3 Scoring rules 5

2.4 Mean absolute error and mean square error 5

3 References **5**

1 MCMC schemes

Posterior inference for both hierarchical and unstructured factor models is performed by the following MCMC algorithms.

1.1 Spatially hierarchical factor model

The full conditional distributions of μ, β^*, θ_0, $f_{ij}(i = 1, \ldots, I, j = 1, \ldots, n_i)$, $\tilde{f}_i(i = 1, \ldots, I)$ and θ are all normal:

(i) μ has variance $C = (C_\mu^{-1} + n\Sigma^{-1})^{-1}$ and mean $C(C_\mu^{-1}\mu_0 + \Sigma^{-1}\sum_{i=1}^{I}\sum_{j=1}^{n_i}(y_{ij} - \beta_i f_{ij}))$.

(ii) β^* has variance $C = (C_0^{-1} + A\tilde{\Sigma}^{-1})^{-1}$ and mean $C(C_0^{-1}\beta_0 + \tilde{\Sigma}^{-1}B)$, where

$$\begin{aligned} A &= \sum_{i=1}^{I}\sum_{j=1}^{n_i} f_{ij}^2 \quad \text{and} \quad B = \sum_{i=1}^{I}\sum_{j=1}^{n_i} f_{ij}(y_{ij}^* - \mu^*) \\ \tilde{\Sigma} &= \text{diag}(\sigma_2^2, \ldots, \sigma_p^2) \\ y_{ij}^* &= (y_{ij2}, \ldots, y_{ijp})' \\ \mu^* &= (\mu_2, \ldots, \mu_p)'. \end{aligned}$$

For C_0 and Σ diagonal,

$$\begin{aligned} \beta_k &\sim N\left(V_k(c_k^{-1}\beta_{0k} + B_k\sigma_k^{-2}), V_k\right) \\ B_k &= \sum_{i=1}^{I}\sum_{j=1}^{n_i} f_{ij}(y_{ijk} - \mu_k) \\ V_k &= (c_k^{-1} + \sum_{i=1}^{I}\sum_{j=1}^{n_i} f_{ij}^2/\sigma_k^2)^{-1} \end{aligned}$$

for $k = 2, \ldots, p$.

(iii) θ_0 has variance $V = (V_0^{-1} + \delta^2 1_I' H(\lambda)^{-1} 1_I)^{-1}$ and mean $V(t_0 V_0^{-1} + \delta^2 1_I' H(\lambda)^{-1}\theta)$.

(iv) f_{ij} has variance $V_i = (\omega_i^{-1} + \beta'\Sigma^{-1}\beta)^{-1}$ and mean $V_i(\omega_i^{-1}(\theta_i + \tilde{f}_{ij}) + \beta'\Sigma^{-1}(y_{ij} - \mu))$.

(v) $\tilde{f}_i$ has variance $V_i = (\omega_i^{-1} I_{n_i} + \tau_i^{-2} P_i^{-1})^{-1}$ and mean $V_i\omega_i^{-1}(f_i - 1_{n_i}\theta_i)$.

(vi) θ has variance $\Sigma_\theta = (\delta^{-2}H(\lambda)^{-1} + X'\Omega^{-1}X)^{-1}$ and mean $\Sigma_\theta(\delta^{-2}H(\lambda)^{-1}1_I\theta_0 + X'\Omega^{-1}F)$, where $X = \text{diag}(1_{n_1}, \ldots, 1_{n_I})$, $\Omega = \text{diag}(\omega_1 I_{n_1}, \ldots, \omega_I I_{n_I})$ and $F = (f_1', \ldots, f_I')' - (\tilde{f}_1', \ldots, \tilde{f}_I')'$.

The full conditional distributions of $\sigma_k^2 (k = 1, \ldots, p)$, $\omega_i^2, \tau_i^2 (i = 1, \ldots, I)$ and δ^2 are all inverse gammas:

(i) σ_k^2 has parameters $a_k + n/2$ and $b_k + \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ijk} - \mu_k - \beta_k f_{ij})^2/2$.

(ii) ω_i^2 has parameters $g_i + n_i/2$ and $h_i + (f_i - 1_{n_i}\theta_i - \tilde{f}_i)'(f_i - 1_{n_i}\theta_i - \tilde{f}_i)/2$.

(iii) τ_i^2 has parameters $c_i + n_i/2$ and $d_i + \tilde{f}_i'(D_i - W_i)\tilde{f}_i/2$.

(iv) δ^2 has parameters $I/2 + a - 1$ and $(\theta - \theta_0 1_I)' H(\lambda)^{-1} (\theta - \theta_0 1_I)/2$.

When sampling λ_1, the decay parameter of our Matérn covariance function, a Metropolis step is performed since its posterior full conditional,

$$\pi(\lambda_1) \propto \lambda_1^{-3} \mathrm{e}^{-h/\lambda_1} |H(\lambda)|^{-1/2} \exp\{-0.5\delta^{-2}(\theta - \theta_0 1_I)' H(\lambda)^{-1} (\theta - \theta_0 1_I)\},$$

is of unknown form. A candidate draw λ_1^* is sampled from a log-normal distribution with location $\log \lambda_1$ and scale Δ_{λ_1} and accepted with probability $\alpha = \min\{1, \pi(\lambda_1^*)\lambda_1^*/\pi(\lambda_1)\lambda_1\}$. We used the algorithm proposed by Rue (2001) and the library `GMRFLib` (see Rue, Follestad, Wist and Martino, 2007, for more details) that rearrange the numbering of the census tracts in order to make P^{-1} as close to a block-diagonal matrix as possible.

1.2 Unstructured hierarchical factor model

Here we present the posterior full conditional distributions of the parameters of the benchmark unstructured hierarchical factor model used as benchmark model in the paper. The full conditional distributions of μ, β^*, θ_0, $f_{ij} (i = 1, \ldots, I, j = 1, \ldots, n_i)$ and θ are all normals.

(i) μ_i has variance $C = (C_\mu^{-1} + n\Sigma^{-1})^{-1}$ and mean $C(C_\mu^{-1}\mu_0 + \Sigma^{-1} \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ij} - \beta f_{ij}))$.

(ii) β^* has variance $C = (C_0^{-1} + A\tilde{\Sigma}^{-1})^{-1}$ and mean $C(C_0^{-1}\beta_0 + \tilde{\Sigma}^{-1}B)$.

(iii) θ_0 has variance $V = (V_0^{-1} + \delta^{-2} I_I)^{-1}$ and mean $V(t_0 V_0^{-1} + \delta^{-2} 1_I' \theta)$.

(iv) f_{ij} has variance $V_i = (\omega_i^{-1} + \beta' \Sigma^{-1} \beta)^{-1}$ and mean $V_i(\omega_i^{-1}\theta_i + \beta_i' \Sigma^{-1}(y_{ij} - \mu))$.

(v) θ has variance $V = (\delta^{-2} I_I + X' \Omega^{-1} X)^{-1}$ and mean $V(\delta^{-2} I_I 1_I \theta_0 + X' \Omega^{-1} F)$.

The full conditional distributions of $\sigma_k^2 (k = 1, \ldots, p)$, $\omega_i^2 (i = 1, \ldots, I)$ and δ^2 are all inverse gammas.

(i) σ_k^2 has parameters $a_k + n/2$ and $b_k + \sum_{i=1}^{I} \sum_{j=1}^{n_i} (y_{ijk} - \mu_k - \beta_k f_{ij})^2/2$.

(ii) ω_i^2 has parameters $g_i + n_i/2$ and $h_i + (f_i - 1_{n_i}\theta_i)'(f_i - 1_{n_i}\theta_i)/2$.

(iii) δ^2 has parameters $I/2 + a - 1$ and $(\theta - \theta_0 1_I)'(\theta - \theta_0 1_I)/2$.

2 Model Selection

In this section we briefly discuss the model comparison criteria we use to compare the models fitted in the paper.

2.1 Expected posterior deviation

The expected posterior deviation (EPD) criterion (Gelfand and Ghosh, 1998) is defined by

$$EPD = G + (\alpha/(\alpha+1))P,$$

where

$$\begin{aligned} G &= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{p}[E(y_{rep,ijk}|y) - y_{ijk}]^2 \\ P &= \sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{p}\text{Var}(y_{rep,ijk}|y), \end{aligned}$$

with $y_{rep,ijk}$ an observation from the posterior predictive distribution. The constant α is used to calibrate the predictive variance P. For example, if $\alpha = 0$, model choice is based purely on goodness of fit and, if $\alpha = \infty$, G and P are weighted equally. In this paper, we assumed the latter. Small EPD is better.

2.2 Deviance information criterion

Spiegelhalter, Best, Carlin and van der Linde (2002) introduce the deviance information criterion (DIC) for model comparison. Let $\Omega = (\Theta, f, \theta, \tilde{f})$ be the vetor of unknown parameters then, the Bayesian deviance is defined as $D(\Omega) = -2\log p(y|\Omega)$ and the DIC is defined by

$$DIC = \bar{D} + p_D,$$

where

$$\bar{D} = E[D(\Omega|y)]$$

is the posterior mean of the deviance and

$$p_D = E[D(\Omega|y)] - D(E[\Omega|y])$$

is the effective number of parameters. Both $\bar{D}$ and p_D can be estimated using the available MCMC output. As a rule of thumb, small DIC is better.

2.3 Scoring rules

Gneiting, Balabdaoui and Raftery (2007) consider scoring rules for assessing the quality of probabilistic forecasts. These measures address calibration and sharpness simultaneously. In particularly, we consider the continuous ranked probability score (CRPS). For each y_{ijk}, the CRPS can be expressed as

$$CRPS(y_{ijk}) = E|y_{rep,ijk} - y_{ijk}| - \frac{1}{2}E|y_{rep,ijk} - \tilde{y}_{rep,ijk}|.$$

In this case, $y_{rep,ijk}$ and $\tilde{y}_{rep,ijk}$ are independent replicates from the posterior predictive distribution. $E|y_{rep,ijk} - y_{ijk}|$ and $E|y_{ijk} - \tilde{y}_{rep,ijk}|$ can be approximated by

$$L^{-1}\sum_{b} l = 1^{L}|y^{(l)}_{rep,ijk} - y_{ijk}| \quad \text{and} \quad L^{-1}\sum_{l=1}^{L}|y^{(l)}_{rep,ijk} - \tilde{y}^{(l)}_{rep,ijk}|,$$

respectively, where, $y^{(l)}_{rep,ijk}$ and $\tilde{y}^{(l)}_{rep,ijk}$ denote replicates from the distribution $p(y_{ijk}|\Omega^{(l)})$ based on the l-th MCMC iterates for $l = 1, \ldots, L$ (see Gschlößl, 2006; Gschlößl and Czado, 2008, for details). Finally, the mean score is given by

$$CRPS = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{p} CRPS(y_{ijk}),$$

for $n = p\sum_{i=1}^{I} n_i$. Small CRPS is better.

2.4 Mean absolute error and mean square error

Standard measures of goodness of fit were also entertained in this study for comparison purposes. They are the mean square errors (MSE) and the mean absolute errors (MAE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{p}(y_{ijk} - \hat{y}_{ijk})^2 \quad \text{and} \quad \text{MAE} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{n_i}\sum_{k=1}^{p}|y_{ijk} - \hat{y}_{ijk}|$$

where $n = p\sum_{i=1}^{I} n_i$ and $\hat{y}_{ijk} = L^{-1}\sum_{l=1}^{L}(\mu_k^{(l)} + \beta_k^{(l)} f_{ij}^{(l)})$ is a Monte Carlo estimate of $E(y_{ijk})$ across L draws. Small MSE and MSE are better.

3 References

- Gschlößl, S. (2006). Hierarchical Bayesian spatial regression models with applications to non-life insurance. Unpublished PhD thesis, Munich University of Technology.

- Gschlößl, S. and Czado, C. (2008). Modelling count data with overdispersion and spatial effects. *Statistical Papers*, **49**, 531-552.
- Gelfand, A. and Ghosh, S. (1998). Model choice: a minimum posterior predictive loss aproach. *Biometrika*, **85**, 1-11.
- Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society, Series B*, **69**, 243-268.
- Rue, H. (2001). Fast sampling of Gaussian Markov random fields. *Journal of the Royal Statistical Society, Series B*, **63**, 325-338.
- Rue, H., Follestad, T., Wist, H. T., and Martino, S. (2007). GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. Technical report, Department of Mathematical Sciences, The Norwegian Institute of Technology.
- Spiegelhalter, D., Best, N., Carlin, B., and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B*, **64**, 583-640.