Statistics in Environmental Research (BUC Workshop Series) II
Problem sheet - WinBUGS

<u>Website</u>: `http://www.stat.ubc.ca/~gavin/STEPIBookNewStyle/`

1. In this question we will carry out Bayesian disease mapping for Ohio lung cancer mortality data from 1988 using `WinBUGS`.

   (a) Code up the Poisson-Gamma model in `WinBUGS` and then analyze the Ohio data, using the expected numbers standardized for age, race and gender, and the priors $p(\beta_0) \propto 1$ and $\alpha \sim \text{Ga}(1, 1)$.

   The `R` command `dput()` is useful for obtaining the data in `WinBUGS` format.

   (b) Extract the posterior means and map these in `R`.

   (c) Calculate the posterior probabilities that the relative risk in each area exceeds 1.2. Extract these probabilities and map these in `R`.

   (d) Now code up the Poisson-Lognormal model, with $p(\beta_0) \propto 1$ and $\sigma_v^{-2} \sim \text{Ga}(1, 0.0260)$. Obtain posterior mean estimates and compare under the Poisson-Gamma model.

2. In this question we will carry out Bayesian disease mapping for Ohio lung cancer mortality data from 1988 using spatial models within `WinBUGS`.

   We analyze the Ohio data using the joint and ICAR conditional models discussed in class. For the joint model assume prior specifications:

$$\begin{aligned}
\tau_T &\sim \text{Ga}(1, 0.026) \\
p &\sim \text{Beta}(1, 1) \\
d_{1/2} &\sim \text{LogNormal}(4.6, 0.42)
\end{aligned}$$

   For the ICAR model assume:

$$\begin{aligned}
\tau_U &\sim \text{Ga}(1, 0.026) \\
\tau_V &\sim \text{Ga}(1, 0.026)
\end{aligned}$$

   (a) Confirm using the `lnprior` function that the prior for $d_{1/2}$ corresponds to a prior for which there is a 5% chance that the correlation drops to a half in less than 50km, and a 95% chance in less than 200km.

   (b) On the website you will find a function `poly.R` which takes polygons created from the `maps()` library and produces a file suitable for reading into `GeoBUGS` where it can be used to draw maps and produce adjacency matrices (for the ICAR model). Carry out this procedure for Ohio.

(c) Produce maps of the posterior mean estimates for the relative risk and for the thresholds under each of the joint and ICAR spatial models.

You will find the county adjacency matrix, assuming a common boundary definition (which you should reproduce above) and the centroids of each county on the class webpage.

3. On the website you will find a reduced set of data from the Alaska study of air pollution in children.

For these data fit using `GeoBUGS` the three-stage model:

(a) Model for the data:

$$\begin{aligned} Y_i|p_i &\sim \text{Bernoulli}(p_i) \\ \log\left(\frac{p_i}{1-p_i}\right) &= \beta_0 + \beta_1 X_i + U_i + V_i \end{aligned}$$

where X_i is the exposure score for child i, $i = 1, ..., n$.

(b) Model for the random effects:

$$\begin{aligned} \boldsymbol{V} = (V_1, ..., V_n) &\sim N(\mathbf{0}, \sigma_v^2 \boldsymbol{I}) \\ \boldsymbol{U} = (U_1, ..., U_n) &\sim N(\mathbf{0}, \sigma_u^2 \boldsymbol{R}) \end{aligned}$$

where we have the exponential model $R_{ij} = \exp(-\phi d_{ij})$.

(c) Hyper-priors:

$$\begin{aligned} \beta_0 &\propto 1 \ \ (\texttt{dflat} \text{ in } \texttt{WinBUGS}) \\ \beta_1 &\propto 1 \ \ (\texttt{dflat} \text{ in } \texttt{WinBUGS}) \\ \sigma_v^{-2} &\sim \text{Ga}(1, 0.0260) \\ \sigma_u^{-2} &\sim \text{Ga}(1, 0.0260) \\ d_{1/2} &\sim \text{LogNormal}(2.30, 0.42) \end{aligned}$$

where $d_{1/2} = \log 2/\phi$.

(d) Report posterior summaries for β_0, β_1, σ_v, σ_u, $d_{1/2}$ and $\sigma_u^2/(\sigma_u^2 + s\sigma_v^2)$.

(e) Carry out an analysis with no spatial random effects U_i to help to answer the question of whether residual spatial dependence is a problem for these data.