DADES EN RETICULES

Generalized linear models and Generalized mixed linear models

Dra. ROSA ABELLANA

DEPARTAMENT SALUT PÚBLICA. UNIVERSITAT DE BARCELONA

The rates of lip cancer in 56 counties in Scotland have been analysed by Clayton and Kaldor (1987) and Breslow and Clayton (1993). The form of the data includes the observed and expected cases (expected numbers based on the population and its age and sex distribution in the county), a covariate measuring the percentage of the population engaged in agriculture, fishing, or forestry, and the "position" of each county expressed as a list of adjacent counties.

County	Observed cases O_i	Expected cases E_i	Percentage in agric. x_i	SMR	Adjacent counties
1	9	1.4	16	652.2	5,9,11,19
2	39	8.7	16	450.3	7,10
...	...	...	...	...	...
56	0	1.8	10	0.0	18,24,30,33,45,55

Poisson regression using R

```
> summary(results.lips)

Call:
glm(formula = O ~ X, family = poisson, data = data.lips, offset =
log(E))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.54227    0.06952   -7.80 6.21e-15 ***
X            0.73732    0.05956   12.38  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 380.73  on 55  degrees of freedom
Residual deviance: 238.62  on 54  degrees of freedom
AIC: 450.6

Number of Fisher Scoring iterations: 5

> deviance(results.lips) /54
[1] 4.418903
```

Dades

```
list(N = 56,
O   = c(    9,   39,   11,    9,   15,    8,   26,    7,    6,   20,
           13,    5,    3,    8,   17,    9,    2,    7,    9,    7,
           16,   31,   11,    7,   19,   15,    7,   10,   16,   11,
            5,    3,    7,    8,   11,    9,   11,    8,    6,    4,
           10,    8,    2,    6,   19,    3,    2,    3,   28,    6,
            1,    1,    1,    1,    0,    0),
E = c( 1.4, 8.7, 3.0, 2.5, 4.3, 2.4, 8.1, 2.3, 2.0, 6.6,
       4.4, 1.8, 1.1, 3.3, 7.8, 4.6, 1.1, 4.2, 5.5, 4.4,
      10.5,22.7, 8.8, 5.6,15.5,12.5, 6.0, 9.0,14.4,10.2,
       4.8, 2.9, 7.0, 8.5,12.3,10.1,12.7, 9.4, 7.2, 5.3,
      18.8,15.8, 4.3,14.6,50.7, 8.2, 5.6, 9.3,88.7,19.6,
       3.4, 3.6, 5.7, 7.0, 4.2, 1.8),
X = c(16,16,10,24,10,24,10, 7, 7,16,
       7,16,10,24, 7,16,10, 7, 7,10,
       7,16,10, 7, 1, 1, 7, 7,10,10,
       7,24,10, 7, 7, 0,10, 1,16, 0,
       1,16,16, 0, 1, 7, 1, 1, 0, 1,
       1, 0, 1, 1,16,10))
```

Model

```
model {

# Likelihood
for (i in 1 : N) {
     O[i]  ~ dpois(mu[i])
      log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + b[i]
      RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + b[i])        }
     # CAR prior distribution for random effects:
    b[1:N] ~ car.normal(adj[], weights[], num[], tau)
     for(k in 1:sumNumNeigh) {
     weights[k] <- 1
}

# Other priors:
     alpha0  ~ dflat()
     alpha1 ~ dnorm(0.0, 1.0E-5)

     tau  ~ dgamma(0.5, 0.0005)   # prior on precision

     sigma <- sqrt(1 / tau)    # standard deviation
}
```

Valors inicials

```
list(tau = 1, alpha0 = 0, alpha1 = 0,
b=c(0,0,0,0,0,NA,0,NA,0,0,
   NA,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0,0,0,0,0,
   0,0,0,0,0,0))
```

The inits can be generate automatically

```
> scotland.moran.mc

        Monte-Carlo simulation of Moran I

data:  nc_scotland$residuals
weights: w.scotland
number of simulations + 1: 101

statistic = 0.33094, observed rank = 101, p-value = 0.009901
alternative hypothesis: greater
```

Code Winbugs heterogeneity model

```
model {
for (i in 1 : I) {
O[i] ~ dpois(mu[i])

log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + h[i]
    RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + h[i])  # Area-specific adjusted relative risk
      h[i] ~ dnorm(0, tau.h)                    # unstructured random effects
      resid[i]<-exp(h[i])
      pp.resid[i]<-step(resid[i]-1)
}

# Other priors:
alpha0 ~ dflat()                    # flat uniform prior on overall intercept
alpha1 ~ dnorm(0.0, 1.0E-5)
tau.h~gamma(0.001, 0.001)
sigma.h<-1/sqrt(tau.h)

..............................
}
```

Gaussian Conditional Autoregression (CAR) models

Between-area covariance matrix:

$$v\mathbf{B} = v(\mathbf{I} - \gamma\, \mathbf{C})^{-1}\, \mathbf{M}$$

where

$\mathbf{I}$	=	N x N identity matrix
$\mathbf{M}$	=	N x N diagonal matrix, with elements M_{ii} proportional to the conditional variance of $S_i \mid S_j$
$\mathbf{C}$	=	N x N weight matrix, with elements C_{ij} reflecting spatial association between areas i and j
γ	=	controls overall strength of spatial dependence

Intrinsic CAR model. Weighted standardized by rows

Distribution of $\theta|\theta_{-i}$ (i=1,...,n) is a Normal with:

$$E\left(\theta_i \mid \theta_{-i}\right) = \mu_i + \sum_j \frac{w_{ij}}{w_{i+}}\left(\theta_j - \mu_j\right)$$

$$\mathrm{Var}\left(\theta_i \mid \theta_{-i}\right) = \frac{\sigma^2}{w_{i+}}$$

$$w_{ij} = \begin{cases} 1 & \text{if region i and j are neighbors} \\ 0 & \text{otherwise} \end{cases}$$

In WinBUGS is defined by:

car.normal(adj[],weights[],num[],tau)

Arguments ***car.normal()*** distribució

adj[]: A vector listing the ID numbers of the adjacent areas for each area

```
adj = c(
19, 9, 5,
10, 7,
12,
28, 20, 18,
19, 12, 1,

17, 16, 13, 10, 2,
................
```

weights[ ]: A vector the same length as ***adj[]*** giving *unnormalised* weights associated with each pair of areas.

num[ ]: A vector of length N (the total number of areas) giving the number of neighbours n_i for each area.

```
num = c(3, 2, 1, 3, 3, 0, 5, 0, 5, 4,
0, 2, 3, 3, 2, 6, 6, 6, 5, 3,.......
```

tau: A scalar argument representing the precision (inverse variance) parameter of the Gaussian CAR prior

num i **adj** can be created with the option **adj matrix** of GeoBUGS **Adjacency Tool**.

Weights:

For (j in 1:**sumNum Nneigh**){weights[]<-1}

On **sumNumNneigh**: Total number of neighbours

car.proper

The proper Gaussian CAR prior distribution is specified using the distribution ***car.proper*** for the *vector* of random variables **S** = (S_1, ..., S_N). The syntax for this distributions is as follows:

S[1:N] ~ car.proper(mu[], C[], adj[], num[], M[], tau, gamma)

where:

mu[] : A vector giving the mean for each area (this can either be entered as data, assigned a prior distribution, or specified deterministically within the model code).
C[] : A vector the same length as ***adj[]*** giving *normalised* weights associated with each pair of areas
adj[] : A vector listing the ID numbers of the adjacent areas for each area
num[] : A vector of length N (the total number of areas) giving the number of neighbours ni for each area.
M[] : A vector of length N giving the diagonal elements Mii of the conditional variance matrix
tau : A scalar parameter representing the overall precision (inverse variance) parameter.

gamma : A scalar parameter representing the overall degree of spatial dependence. This parameter is constrained to lie between bounds given by the inverse of the minimum and maximum eigenvalues of the matrix $\mathbf{M}^{-1/2}$ $\mathbf{C}$ $\mathbf{M}^{-1/2}$ (see appendix). *GeoBUGS 1.1Beta* contains two deterministic functions for calculating these bounds:

```
min.bound(C[], adj[], num[], M[])
max.bound(C[], adj[], num[], M[])
```

where the arguments are the same as the corresponding vectors used as arguments to the ***car.proper*** distribution.

Codi WinGUBS

```
model
{
  for (i in 1 : N) {
      O[i]  ~ dpois(mu[i])
      log(mu[i]) <- log(E[i]) + alpha0 + alpha1 * X[i]/10 + b[i]
      RR[i] <- exp(alpha0 + alpha1 * X[i]/10 + b[i])
                 # Area-specific relative risk (for maps)
  }

  # CAR prior distribution for random effects:
  b[1:N] ~ car.normal(adj[], weights[], num[], tau.b)
  for(k in 1:sumNumNeigh) {
      weights[k] <- 1
  }

  # Other priors:
  alpha0  ~ dflat()
  alpha1 ~ dnorm(0.0, 1.0E-5)
  tau.b  ~ dgamma(0.5, 0.0005)          # prior on precision
  sigma.b <- sqrt(1 / tau.b)                # standard deviation
}
```

Convolution prior

To allow greater flexibility, Besag, York and Mollie (1991) recommend combining the intrinsic CAR prior and an exchangeable normal prior:

$$
\begin{aligned}
\theta_i &= S_i + H_i \\
H_i &\sim \text{Normal}(0, v^2) \\
S_i | S_{-i} &\sim \text{Normal}(m_i, s_i^2) \\
\text{where} \quad & m_i = \frac{\sum_j w_{ij} S_j}{w_{i+}}; \quad s_i^2 = \frac{s^2}{w_{i+}}
\end{aligned}
$$

- Total variation in reflects a combination of spatial dependence and unstructured heterogeneity
– The data determine the relative contribution of each component

Inferència sobre la contribució de l'efecte espaial i d'heterogeneïtat en el model de convolució

υ^2(variança no estructurada) i s^2 (variança espaial) no són comparables directament:

- υ^2 reflecteix la variabilitat marginal de l'efecte aleatori no estructurat

- s^2/n_i reflecteix la variança conditional de l'efecte aleatori de l'area i condicionada als valors dels efectes aleatoris dels seus veïns

No hi ha una expressió tancada de la variança marginal entre regions

- Estimar la variança marginal del efectes aleatoris empíricament

$$s^2_{\text{marginal}} = \sum_i (S_i - \overline{S})^2/(I-1)$$

Contribució relativa de espacial respecte la no estructura heterogeneïtat

$$\text{frac}_{\text{spatial}} = s^2_{\text{marginal}}/(s^2_{\text{marginal}} + v^2)$$

$\text{frac}_{\text{spatial}} \rightarrow 1$ espaial domina sobre heterogeneïtat

$\text{frac}_{\text{spatial}} \rightarrow 0$ heterogeneïtat no estructurada domina

Codi WinBUGS for convolution model

```
model {
for (i in 1 : I) {
O[i] ~ dpois(mu[i])
log(mu[i]) <- log(E[i]) + alpha + theta[i]
theta[i] <- H[i] + S[i]
H[i] ~ dnorm(0, v2.inv)                # unstructured random effects
adjustedR[i] <- exp(theta[i])          # Area-specific adjusted relative risk
}
# intrinsic CAR prior on spatial random effects
S[1:I] ~ car.normal(adj[], weights[], num[], s2.inv)
for(k in 1:sumNumNeigh)
        {weights[k] <- 1 }    # adjacency weights for CAR model
# Other priors:
alpha ~ dflat()                        # flat uniform prior on overall intercept
............................
s2.marginal <- sd(S[]) * sd(S[]) # empirical marginal var. of spatial effects

# fraction of total variation in log relative risks due to spatial effects

frac.spatial <- s2.marginal / (s2.marginal + v2)

}
```

Bayesian models: Deviance information criterion (DIC)

Spielgelhalter et al (2002), proposed to compare models:

Deviance Information Criterion DIC= “Goodness of fit” +complexity

Measure goodness of fit using deviance:

$$D(\theta) = -2\log L(Y \mid \theta)$$

Measure complexity of the model using the effective number of parameters

$$\begin{aligned} p_D &= E_{\theta|Y}[D] - D(E_{\theta|Y}[\theta]) \\ &= \overline{D} - D(\overline{\theta}) \end{aligned}$$

DIC is defined as AIC :

$$DIC = \overline{D} + p_D$$

Lower DIC is better