

Final Project: Poisson and Negative Binomial Regression

This dataset refers to outbreaks of tuberculosis in dairy and beef cattle, cervids and bison in Canada between 1985 and 1994. There are $n=134$ observations.

The predictors are type of animal, a categorical variable with four levels (1=Dairy, 2=Cervid, 3=Beef, 4=Other), male sex, and age, also categorical, with levels '0-12 months' = 1, '12-24 months' = 2, '>24 months' = 3. Other inputs are an offset *adar*, namely animal days at risk.

The response is called *reactor*, for count of TB reactors. The data are called *tb.txt* in the supplied materials.

Fit a Poisson model with predictors *type*, *sex* and *age*, including the following features:

a) *adar* as an offset, as in

```
mu[i] <- adar[i]*nu[i]
log(nu[i]) <- beta0+...
```

b) fixed effects with a corner constraint to represent the impact of type of animal

b) fixed effects with a corner constraint to represent the impact of age.

One way to handle categorical predictors are illustrated by the drug dependency analysis (section 3.3.3). Another (equally correct) way would be something like

```
log(nu[i]) <- beta0+beta.type[type[i]]...
```

with the prior on `beta.type[1:4]` setting `beta.type[1]` to zero, and

assigning the remaining `beta.type` parameters $N(0,1000)$ priors, as in

```
beta.type[1] <- 0; for (j in 2:4) {beta.type[j] ~ dnorm(0,0.001)}
```

Note that in setting initial values, one then needs to set the constrained value (which is no longer an unknown parameter) as NA, so the initial values could be `beta.type=c(NA,0,0,0)`.

Include commands to monitor the deviance contribution of each observation, for example

```
dv[i] <- 2*(reactors[i]*log(reactors[i]/mu[i])-(reactors[i]-mu[i])),  
and also the total deviance.
```

Questions:

Q1: Is the posterior mean deviance (a) above 300 or (b) below 300.

Q2: Assuming dairy cattle is the reference category for the type effects, is the Cervid (type 2) effect significant (i.e. has a 95 credible interval entirely positive), or not significant (95% interval straddles zero). Note that to monitor `beta.type` it is better (for the purposes of obtaining Gelman-Rubin PSRF values) to use the form

```
R <- coda.samples(M,c("beta.type[2:4]",...),n.iter=...)
```

Q3: Modify the Poisson model to a negative binomial model, for example as in

```
for (i in 1:134) { reactors[i] ~ dnegbin(p[i],theta)  
  mu[i] <- nu[i]*adar[i]  
  p[i] <- theta/(theta+mu[i])  
  log(nu[i]) <-
```

etc. Use a $U(0,10)$ prior on the shape parameter θ . How many of the three animal type effects are now significant (with 95% credible intervals either entirely +ve, or entirely -ve).

Q4 Is the posterior mean for the shape parameter θ in the NB model (a) above 0.5 or (b) below 0.5.