BAYESIAN STATISTICAL MODELS - University of Milano-Bicocca

Case study - grouses and ants

Lecturer: Riccardo Corradin

First, we have a set of 403 observations, where each observation correspond to a particular grouse. You can find the data in the file *grouses.csv*, available at the module page. Specifically, for each observed animal we measure

- Height, above sea level (meters), denoted by X , where the animal lives, continuous variable.
- Year (three different years), denoted by G , where G = 1 is 1994, G = 2 is 1995 and G = 3 is 1996.
- Number of ticks Y on the heads of the red grouse, which plays the role of response variable.

We are interested study the number of ticks on the grouses' head, as function of the other observed quantities. Specifically, we want to consider a Poisson regression model including a common intercept and random effects for the last two years in the study. The model specification is then given by

$$
\begin{aligned}
Y_{i,j} \mid \mu_{i,j} &\sim Poi(\mu_{i,j}), \qquad i = 1, \ldots, n_j,\ j = 1, 2, 3, \\
\mu_{i,j} &= \mathrm{e}^{\beta_1 + \beta_2 x_{i,j} + \gamma_j}, \qquad i = 1, \ldots, n_j,\ j = 1, 2, 3, \\
\boldsymbol{\beta} &\sim N(\mathbf{0}, \Sigma_0), \\
\gamma_1 = 0, &\qquad \gamma_j \sim N(0, \sigma_0^2),\ j = 2, 3
\end{aligned}
$$

where $Y_{i,j}$ denotes the ith observation of the jth year, $\mu_{i,j}$ the ith expectation of the jth year, (β_1, β_2) are our fixed effects and γ_j, $j = 2, 3$, our random effects.

a) Implement the model in STAN. Think wisely about how to include the random effects.

b) Choose specific values for the dispersion parameters of the prior distributions, Σ_0 and σ_0^2, and produce a sample of size $5\,000$ from the posterior distribution of interest, of which $2\,000$ are burnin values.

c) Check if the random coefficients are significantly different from 0, by looking at their credible intervals.

d) Produce a plot that shows the model behavior for each separated year. Ideally, you should show a plot with three line, one for each year, that corresponds to the evolution of the Poisson distribution expectation as far as the covariate is moving on its support. Hint: restrict the values you are considering of the covariates on the observed range.

e) Produce a second model estimate without the random effects. Test if the two models are significantly different from each other.

Now, we consider a set of 60 observations of an ecological experiment. Each observation corresponds to a specific tree. The trees are divided into High and Low altitude trees, defining two macro-groups. The experiment wants to study the amount of ants that are populating each specific tree. For each spot we observe:

- Trap days, different lengths of time surveying each tree, denoted by X.
- Elevation, G . If G = 1, then the spot is High, if G = 2 then the spot is Low.
- Valley, three different valleys for each elevation, denoted by SG, nested within each elevation.
- Summer precipitation (avg), assuming a random effect, denoted by Z1.
- Summer temperature (avg), assuming a random effect, denoted by Z2.
- Number of ants, assuming a random effect, denoted by Y.

We assume a priori a Poisson distribution of the data, having a model of the form

$$
\begin{aligned}
Y_{i,j,r} \mid \mu_{i,j,r} &\sim Poi(\mu_{i,j,r}), \qquad i = 1, \ldots, n_j,\ j = 1, 2,\ r = 1, 2, 3, \\
\mu_{i,j,r} &= \exp\{\beta x_{i,j,r} + \theta_j + \gamma_{j,1} z_{1,i,j,r} + \gamma_{j,2} z_{2,i,j,r} + \alpha_{j,r}\}, \qquad i, j, r = \ldots, \\
\beta &\sim N(0, \sigma_0^2), \\
\theta_j &\sim N(0, \eta_0^2), \qquad j = 1, 2, \\
\gamma_{j,1}, \gamma_{j,2} &\sim N(0, \tau_0^2), \qquad j = 1, 2, \\
\alpha_{j,r} &\sim N(0, \lambda_0^2)
\end{aligned}
$$

where β is the fixed effect, associated to the trap days, θ_js are random intercepts specific for each elevation, $\alpha_{j,r}$ are nested random intercepts specific for each valley, $\gamma_{j,1}$ and $\gamma_{j,2}$ are random effects associated to summer precipitation and temperature.

a) Implement the model in STAN.

b) Choose specific values for the dispersion parameters of the prior distributions, σ_0^2, η_0^2, τ_0^2, λ_0^2. Produce a sample of size $5\,000$ from the posterior distribution of interest, of which $2\,000$ are burnin values. Looking at the traceplots, think about the model specification and possibly discard some random intercepts.

c) Check if the random intercepts, θ_js and $\alpha_{j,r}$s, are significantly different from 0, by looking at their credible intervals.

d) Produce a second model estimate without the random effects, only having a common intercept term and fixed effects for X, Z1 and Z2. Test if the two models are significantly different from each other.