ST740 – Assignment 4 – Due 11/9

The homework is due on 11/9 and should be submitted on moodle. **Recall, your final homework percentage will the average after dropping your lowest of the four scores, so if you are happy with your grade after A1-A3 then you do not have to complete this assignment**. Show work for questions that say "derive" and include code used to produce the results. You many work in groups, but write up separate solutions.

Download the data [1] as below:

```
file <- "https://www4.stat.ncsu.edu/~bjreich/ST740/blgr.csv"
Y    <- read.csv(url(file))[,2:4]
X    <- read.csv(url(file))[,5]
```

The matrix `Y` is an $n_s = 41$ by $n_t = 3$ matrix with Y_{st} being the binary indicator that a Blue Grosbeak was observed in field s at visit t. The vector `X` gives the field size of each of the n_s fields. Let $Z_s = 1$ if Blue Grosbeaks truly occupies field s, $Z_s = 0$ otherwise and assume the model

$$\begin{aligned} Y_{st}|Z_s, p &\overset{indep}{\sim} \text{Bernoulli}(\theta Z_s) \\ \text{Logit}\{\text{Prob}(Z_s = 1|\alpha, \beta)\} &= \alpha + X_s\beta \end{aligned}$$

with prior $\theta \sim \text{Uniform}(0, 1)$ and $\alpha, \beta \sim \text{Normal}(0, c^2)$.

1. Give a one-sentence interpretation of each of the parameters θ, α and β.
2. What are the main assumptions being made?
3. Derive the full conditional distributions of Z_s and θ.
4. Write a Gibbs-within-Metropolis sampler and verify that the chain has converged
5. Summarize the results of your analysis, including (a) a test of whether field size is associated with occupancy and (b) and the posterior probability that all n_s fields are truly occupied.
6. Are the results sensitive your the hyperparameter c? Which value of c would you choose for the final analysis and why?

[1] https://sites.google.com/site/asrworkshop/home/schedule/r-occupancy-1