NRES 779: Bayesian Hierarchical Modeling in Natural Resources

Lab 08: JAGS Problems

Contents

I Motivation

JAGS allows you to implement models of high dimension once you master its syntax and logic. It is a great tool for ecological analysis. The problems that follow challenge you to:

- Write joint distributions as a basis for writing JAGS code.
- Write JAGS code to compute posterior distributions of derived quantities.
- Plot model output in revealing ways.
- Understand the effect of vague priors on parameters on predictions of non-linear models.

II Derived quantities with the logistic

One of the most useful features of MCMC is its equivariance property which means that any quantity that is a function of a random variable in the MCMC algorithm becomes a random variable. Consider two quantities of interest that are functions of our estimates of the random variables r and K from our logistic regression model we developed in the JAGS primer.

- The maximum rate of population growth, $\frac{K}{2}$
- The rate of population growth, $\frac{dN}{dt} = r(1 - \frac{N}{K})$

1. Estimate the posterior distribution of the intrinsic growth (r) rate and the maximum rate of growth ($K/2$) and plot their densities. You may use the work you have already done in the Primer to speed this along.
2. Plot the median population growth rate and 95% credible intervals as a function of N (maybe $0, 10, 20, \ldots, 1100$?). What does this curve tell you about the difficulty of sustaining harvest in populations?
3. What is the probability that the intrinsic rate of increase (r) exceeds 0.22? What is the probability that r falls between 0.18 and 0.22?

   Some hints:

   - Include expressions for each derived quantity in your JAGS code.
   - You will need to give JAGS a vector of N values to plot $\frac{dN}{dt}$ vs. N.
   - Use a JAGS object for plotting the rate of population growth.
   - Look into using the `ecdf()` function on a JAGS object. It is covered in the JAGS primer.

III Lizards on islands

This problem is courtesy of McCarthy (2007). Polis et al. (1998) analyzed the probability of occupancy of islands p by lizards as a function of the ratio of the islands' perimeter to area ratios. The data from this investigation are available in the data frame `IslandLizards`. The response data, as you will see, are 0 or 1; 0 if there were no lizards found on the island, 1 if there were 1 or more lizards observed. You are heroically assuming that if you fail to find a lizard, none are present on the island.

4. Construct a simple Bayesian model that represents the probability of occupancy as $\text{logit}(p_i) = \beta_0 + \beta_1 x_i$, where x_i is the perimeter to area ratio of the ith island. So, now that you have the deterministic model, the challenge is to choose the proper likelihood to link the data to the model. How do the data arise? What likelihood function is needed to represent the data?
5. Write out the expression for the posterior distribution of the parameters given the data, as we have learned how to do in lecture. Use the posterior distribution as a basis for JAGS code needed to estimate the posterior distribution of β_0 and β_1. Assume vague priors on the intercept and slope, e.g., $\beta_0 \sim \text{Normal}(0, 10000)$, $\beta_1 \sim \text{Normal}(0, 10000)$.
6. Using JAGS, run MCMC for three chains. Selecting initial conditions can be a bit tricky with the type of likelihood you will use here. You may get the message "Error in jags.model("IslandsJags.R", data = data, inits, n.chains = length(inits), : Error in node y[4] Observed node inconsistent with unobserved parents at initialization". To overcome this, try the following: .

   - Standardize the the perimeter to area ratio covariate using the `scale` function in R. Review the Bayesian Regression lecture for details on standardizing the data. You want the default arguments for center and scale in this function.
   - Choose initial values for β_0 and β_1 so that $\text{logit}^{-1}(a + b\text{standardized}(x_i))$ is between 0.01 and 0.99.
7. Do a plot of the posterior density and the trace of the chain using the plot(zm). Does the trace indicate convergence? How can you tell? Use Gelman and Heidel diagnostics to check for convergence.
8. Plot the data as points. Overlay a line plot of the median and 95% credible intervals of the predicted probability of occurrence as a function of island perimeter to area ratios ranging from 1–60. Hint–create a vector of 1–60 in R, and use it as x values for an equation making predictions in your JAGS code. Use a JAGS object for plotting. This makes a nice smooth curve. (The curve is jumpy if you simply plot the predictions at the island perimeter to area data points.)

9. Assume you are interested in 2 islands, one that has a perimeter to area ratio of 10, the other that has a perimeter to area ratio of 20. What is the 95% credible interval on the difference in the probability of occupancy of the two islands based on the analysis you did above? Remember that the data are standardized when you do this computation.

10. What fundamentally important source of error are we sweeping under the rug in all of these fancy calculations? What are the consequences of failing to consider this error for our estimates? Do you have some ideas about how we might cope with this problem?

    The priors you chose above were vague for the intercept and slope in the logistic regression but they were not vague for p_i . This is generally true for the output of nonlinear functions like the inverse logit (Lunn et al., 2012; Seaman et al., 2012), so you need to be careful about inference on the output of these non-linear function. For an explanation of this particular case (a logistic regression), see Hobbs and Hooten (2015) section 5.4.1. The solution is to explore the effect of different values for priors on the shape of a the "prior for quantities that are non-linear functions of model parameters, as demonstrated in the following exercise.

11. Write a function that takes an argument for the variance σ^2 . The function should 1) simulate 10000 draws from a normal distribution with mean 0 and variance σ^2 representing a prior on β_0 , remembering, of course, that the argument to `rnorm` is the standard deviation. 2) Plot histograms of the draws for β_0 . 3) Plot a histogram of the inverse logit of the random draws, representing a "prior" on p at the mean of x (i.e., where the scaled value of $x = 0$). Plotting these in side by side panels will facilitate comparison. Use your function to explore the effect of different variances ranging from 1 to 10000 on the priors for β_0 and p . Find a value for the variance that produces a flat "prior" on p.

12. Rerun your analysis using priors on the coefficients that are vague for inference on p based on what you learned in Hobbs and Hooten section 5.4.1 and in exercise 8. (Be careful to convert variances to precision) Plot the probability of occupancy as a function of perimeter to area ratio using these priors and compare with the plot you obtained in exercise 5, above. You will see that the means of the p_i changes and uncertainty about p_i increases when you use appropriately vague priors for p .

IV Guidance

There is conflict between priors that are vague for the parameters and vague for the predictions of the model. If your primary inference is on p then you want to choose values for the priors on β_0 and β_1 that are minimally informative for p . The simulation exercise above shows a way to do that. However, what if you need inference on β_0 , β_1 , and p ? There are two possibilities. First, get more data so that the influence of the prior becomes small. Second, use informative priors on the coefficients, even weakly informative ones. For example, you know that the slope should be negative and you know something about the probability of occupancy when islands are large. Centering the slope on a negative value rather than 0 makes sense because we know from many other studies that the probability of occupancy goes down as islands get smaller. Moreover, you could center the prior on the intercept on 3 using the reasoning that large islands are almost certainly occupied (when intercept = 3, p = 0.95 at PA = 0). Centering the priors on reasonable values (rather than 0) will make the results more precise and far less sensitive to the variance (or precision) chosen for the prior.

Informative priors, even weakly informative ones, are helpful in many ways. We should use them.

V References

McCarthy, M.A. 2007. Bayesian Methods for Ecology. Cambridge University Press, Cambridge, United Kingdom.

Polis, G.A., S.D. Hurd, C.T. Jackson, and F. Sanchez-Pinero, 1998. Multifactor population limitation: Variable spatial and temporal control of spiders on gulf of California islands. Ecology 79:490–502.

Seaman III, J.W., J.W. Seaman Jr., and J.D. Stamey. 2012. Hidden dangers of specifying noninformative priors. The American Statistician 66, 77-84.

Hobbs, N. T. and M. B. Hooten. 2015. Bayesian models: A statistical primer for ecologists. Princeton University Press, Princeton, New Jersey, USA.

Lunn, D., C. Jackson, N. Best, A. Thomas, and D. Spiegelhalter. 2012. The BUGS book: A practical introduction to Bayesian analysis. CRC Press, Boca Raton, Florida, USA.