BIOL 701 Likelihood Methods in Biology: Homework #7

Due Wednesday, May 11th and May 18th (see below)

Background

You are interested in fecundity of an annual plant (as assessed by the total mass of seeds produced measured in g) as a function of fertilization regimes. You have collected data from several years in eight different fields. Within each field season, you examined plants four different treatments in each field. These treatments correspond to growing the plants with (coded as 1) and without (coded as 0) fertilization with a nitrogen source and a phosphorus source. Thus, the four treatments are: no fertilization, +N, +P, and +N,P. For each treatment$\times$field$\times$year combination, you recorded data for four plants.

You are interested in inferring the effectiveness of different treatments in general (e.g. the expected effect of adding Nitrogen to some unspecified field), as well as learning about which fields have the highest yield.

Consider a model in which each of the following effects contribute additively to the expected mass for an individual:

1. a year effect (centered around 0),
2. a field-specific expected mass without fertilization (this effect is the same for a field across all years),
3. a field-specific effect of nitrogen fertilization (this effect is the same for a field across all years),
4. a field-specific effect of phosphorus fertilization (this effect is the same for a field across all years)
5. a field-specific effect of adding both N and P (this effect is the same for a field across all years).

Of course, you should also expect some variability around this expected value (not all individuals from the same treatment, field, and year will have exactly the same mass).

Note that you would like to make some general conclusions about the effect of different fertilizer treatments on some hypothetical field. So, you should structure the model so that you can learn about expected effects across fields, while accounting for the fact that their may be variability in the response of a particular field to a particular treatment.

Tasks (still due May 11th)

1. Write down the likelihood for your model.
2. List each parameter and the prior distribution that you have chosen to use for the parameter.

2.5 Derive the formulae the log-likelihoods needed to update each parameter/latent variable. Note that several log-likelihood formulae that you derive will be slight variants on the same theme. In these cases, it is fine to derive these formula once, and then add a like “the update of β will be just like the update for α except all of the A_j terms will become B_j"

Tasks originally assigned to be completed by May 11th, but pushed back to May 18th

2.6 If you did not derive and write down the formulae for the log-likelihood-*ratio* in step 2.5, then do so now.

3 Implement an MCMC algorithm that uses the data in fertilization_data.csv to inference posterior distributions for the parameters. latent_gekko_svl.py.txt may be useful as a template. Email me your implementation.

4 Perform an MCMC simulation. Summarize the evidence that your MCMC run has been conducted for enough iterations (Tracer or CODA may be helpful for this) to generate useful results.

5 Based on your runs, answer the following questions (and give a brief explanation of how you calculate the answers).
   (a) If you could scale the experiment up to an unlimited number of fields, what is the probability that the mean effect of adding nitrogen alone will be an effect of 5g or greater?
   (b) Which field has the largest expected mass without fertilization?
   (c) What is the probability is the probability that field # 6 is the most productive (has the highest expected mass over a large number of years) if both N and P are added?

New tasks (due May 18th)

6 Derive a new proposal algorithm for the error variance in your model. If the current value of the parameter is θ, then your proposal should be capable of proposing new values on the interval $[\sqrt{\theta}, \theta^2]$.

7 Calculate the Hastings ratio for your new proposal.

8 Implement it in the MCMC software.

9 Does the new proposal seem to affect the "mixing" of the MCMC simulation?