

**Homework # 2**, Population models,

Due: 5 pm, Friday, 15 November 2007, to my office (120 Snedecor) or mailbox (in 115 Snedecor)

Reminders: Choose **4 of the 5** problems. You are not expected to do all five. Also, we have not talked about assumptions or diagnostics for any of these models. You are not expected to do any here.

This set of problems doesn't divide clearly into 'data' and 'theory' problems. Please consider problems 4a and 5a as 'data problems' for which I require an executive summary.

## Problem 1.

*Avicennia marina*, the grey mangrove, is an abundant plant on mudflats in southern hemisphere estuaries (see photo linked to class homepage for pictures of *A. resinosa*, a related species in New Zealand). Mangrove forests can be disturbed in two different ways. Small-scale disturbances bring down one or a few mature trees. Large-scale disturbances (storms, disease, or sedimentation) create large areas of open habitat. P. J. Clarke in Australia collected demographic data on mangroves in plots with small-scale disturbances and in plots with large-scale disturbances.

Clarke used 7 stages to describe the demography and constructed transition matrices by following marked plants. The transition matrices describe the growth, survival and fecundity of those plants in habitats with the two types of disturbance. The data are in `avicennia.txt` in the data files section of the class web page. The survival and growth parts of these transition matrices are based on observations of many plants. The fecundity is much more uncertain. Using the limited information that is available about the uncertainty of each element in the transition matrix, I compute the uncertainty in  $\lambda$  as s.e.  $\lambda \approx 0.07$ .

- a) Estimate  $\lambda$  for plants with small scale disturbances. Construct a 95% confidence interval for  $\lambda$  by transforming a 95% ci for  $\log \lambda$ .
- b) Which elements of the transition matrix have the highest elasticity?
- c) Describe whether and by how much the difference in disturbance affects the demography of these plants. Which transitions contribute the most to the difference in  $\lambda$ ?

Problem 2. Miscellaneous analytical issues.

2-a) Sensitivity to changes in multiple vital rates. Many stage-based transition matrices have the following form:

$$\begin{array}{cccc} s_1(1 - g_1) & f_2 & f_3 & f_4 \\ s_1g_1 & s_2(1 - g_2) & 0 & 0 \\ 0 & s_2g_2 & s_3(1 - g_3) & 0 \\ 0 & 0 & s_3g_3 & s_4 \end{array}$$

where  $s_i$  is a stage-specific survival probability and  $g_i$  is a stage-specific conditional probability that a surviving plant will grow into the next stage. For many species, the sensitivity of  $\lambda$  to changes in  $s_i$  or  $g_i$  may be more important than the sensitivities we have computed (i.e.,  $\partial\lambda/\partial a_{ij}$ ). The sensitivities to  $s_i$  or  $g_i$  can be computed from the “usual” sensitivities. Derive the expressions to compute  $\partial\lambda/\partial s_i$  and  $\partial\lambda/\partial g_i$ .

HINT: think about the chain rule for derivatives.

2-b) Analytical Bayesian calculations. My simple illustration of Bayes in class used the single normal sample.

$$\begin{aligned} Y_i &\sim N(\mu, \sigma^2) \\ \mu &\sim N(\eta, \sigma_\eta^2) \end{aligned}$$

where  $\sigma^2$ , the data variance, is known and  $\eta$  and  $\sigma^2$  the prior mean and variance are specified. I provided the posterior distribution in class. Please derive this.

Almost full credit will be given for a derivation of the moments, assuming the posterior distribution is normal.

### Problem 3. Validating population estimates

Graduate student research is the source of a lot of demographic data. However, students tend to collect lots of data for (relatively) short periods of time. They often collect age- or stage-classified information. State Departments of Natural Resources are often required to monitor population sizes of rare or endangered plants and animals. They usually do not have the time to continue detailed demographic monitoring. Commonly, they only track the total population size over time. So, the following history of information about a specific population is not too uncommon:

Year	Type of data collected	Information available
1	Student starts study	Initial counts, i.e. $N_1$ and $\tilde{N}_1$
2	Student continues	$N_2$ and $\tilde{N}_2$
3	Student continues	$N_3$ and $\tilde{N}_3$
4	Student finishes	$N_4$ and $\tilde{N}_4$
5	DNR starts	$N_5$
6	DNR continues	$N_6$
7, 8, 9, 10	DNR continues	$N_7, N_8, N_9,$ and $N_{10}$

where:

$N_i$  is the total number of individuals in the population in year  $i$ , and

$\tilde{N}_i$  is the vector of counts of stage-classified individuals in year  $i$

Imagine the student estimates a pooled (over years 1-4) transition matrix, then uses a deterministic matrix model to estimate  $\hat{\lambda}$ . The DNR has come to you, an eminent statistician, to see if they can validate this estimate. In other words, could they use the data from years 5-10 to see whether the population is continuing to grow or decline at  $\lambda$  per year?

3-a) The DNR tells you that the rare species is an animal and that  $N_i$  and  $\tilde{N}_i$  were estimated using mark-recapture sampling. The sampling variation in these estimates is large. Is it possible to evaluate whether the population continues to grow at  $\hat{\lambda}$ ? Either indicate how you would do this or why it could not be done.

3-b) The DNR tells you that the rare species is a plant and the population grows in a very defined area. So, the DNR believes the population counts ( $N_i$  and  $\tilde{N}_i$ ) are exact. Is it possible to evaluate whether the population continues to grow at  $\hat{\lambda}$ ? Either indicate how you would do this or why it could not be done.

Problem 4. Estimating the transition matrix from observed data. In class, I preceeded the state-space discussion by contrasting two approaches to analyze a time series of abundances. The 'process error only' approach used regression, conditioning on the numbers at the previous time. The 'observational error only' approach compared the observed counts to projections at each time. My illustration had 4 stages; to keep things simple, I will use only 2 stages (juvenile and adult) here. The appropriate transition matrix is:

$$\begin{bmatrix} 0 & f \\ s_j & s_a \end{bmatrix}$$

where:

$s_j$  is survival of juveniles,

$s_a$  is survival of adults, and

$f$  is fecundity, number of survival juveniles per adult

A population has been studied for 20 years; the counts of juveniles and adults are given in abdn2.txt.

4a) Assume that all variation represents process error, so it is appropriate to condition on the number of individuals in the preceding generation. You may assume that process errors are normally distributed with constant variance. Please estimate the unknown coefficients in the transition matrix. Also estimate their standard errors.

4b) You realize that the assumption in 4a of only process error is unreasonable. An appropriate model needs to incorporate both process and observation error. Write out an appropriate model that does this. Provide distributions for any random variables in your model ('good guesses' for an appropriate distribution are sufficient). Define your parameters and variables in terms a biologist would understand.

You do not need to fit this model, but I'll give some extra credit if you do.

4c) If you did have results from 4b, what could you use to evaluate whether the results in 4a were appropriate?

This is a concept (what would you do?) question, so you need to answer it even if you don't fit the model in 4b.

Problem 5 - estimating mutation rates in viruses.

Part a) Viruses mutate very quickly. That's why we have no vaccine against the common cold. PRRS (Porcine Respiratory and Reproductive Syndrome) is a viral disease that causes major problems for pig producers. Researchers would like to estimate the mutation rate in vivo. This problem uses a very simple model that could be used to do this. The simple version of the model is a birth process with random variation in the birth rate. That is new mutations are 'born' at rate  $b$  per unit time, independent of the number of mutations in the virus population. Because of the experimental set up, the birth rate is expected to vary each time unit (process error). The experiment starts with a single clone of virus, so there are no mutations present at time 0. The observed data are counts of observed number of mutations (new genotypes) in a small sample of virus clones. The sampling process is such that the observed number of mutations has a Poisson distribution with  $\lambda_t = 0.1 * N_t$ , where  $N_t$  is the 'true' number of mutations present at time  $t$ .

The model could be written as:

$$\begin{aligned}\log b_t &\sim N(\log b, \sigma_b^2) \\ N_0 &= 0 \\ N_{t+1} &= N_t + b_t \\ Y_{t+1} &\sim Poiss(0.1 * N_t)\end{aligned}$$

The data in prrs.txt are the time series of abundances,  $Y_t$ , for 30 time units. Please estimate the mean and a 95% credible interval for  $b$ .

Part b) The model above ignores one crucial piece of biology. Mutations can 'die', i.e. disappear from the population. In a birth-death process, the death rate  $d$  is constant **per mutation** per time, so the average number of deaths in a population that starts with  $N_i$  mutations is  $dN_i$ . When there are more mutations in the population, more of them (on average) disappear. Again because of the experimental set up, the death rate is expected to vary each time unit (process error) and it is reasonable to assume that the death rate variation and birth variation are independent.

Write out a model, including reasonable choices of prior distributions, for this birth-death process. You may choose an appropriate scale for the process variation. One point that may need care is that the distribution of  $d_t$  is truncated,  $0 \leq d_t \leq 1$ . (At most 100% of the mutations can disappear).

For extra credit, fit this model to the data in prrs2.txt. Report the mean and 95% credible intervals for  $b$  and  $d$  and plot the joint posterior distribution of  $b$  and  $d$ .