STAT 535C - CS 535D ASSIGNEMENT #1, DUE IN CLASS ON FEBRUARY, 23.

Capture-Recapture Experiments

We consider here the problem of estimating an unknown population size, N, based on observations related with this population. This is based on a simple Bayesian model.

1. Assume we capture n individuals from the population of size N. While the population size is the parameter of interest, there is a nuisance parameter, namely the probability $p \in [0,1]$ with which each individual is captured. For this model, we have
$$n \sim \mathcal{B}(N, p)$$
where $\mathcal{B}(\cdot)$ is a binomial distribution.

   Assume we select the following prior
$$\pi(N, p) = \pi(N)\,\pi(p)$$
where
$$\begin{aligned} \pi(p) &= \mathcal{B}e(1,1), \\ \pi(N) &\propto \frac{1}{N}\mathbf{I}_{\mathbb{N}^*}(N) \text{ (improper prior)} \end{aligned}$$
and $\mathcal{B}e(\alpha, \beta)$ denotes the Beta distribution of parameter (α, β) (so $\mathcal{B}e(1,1)$ is the uniform distribution on $[0,1]$).

   Establish analytically the marginal posterior distribution for $\pi(N|\,n)$.

2. Consider now the more realistic scenario where I samples of sizes $n_1, n_2, ..., n_I$ are consecutively drawn, marked and returned to the population. When the probability of any capture in the i^{th} sample is p_i then
$$n_i \sim \mathcal{B}(N, p_i).$$
The following Gordy Lake sunfish data set will be analyzed. It consists of $I = 14$ capture occasions from a population of sunfish.

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14
n_i	10	27	17	7	1	5	6	15	9	18	16	5	7	19
m_i	0	0	0	0	0	0	2	1	5	5	4	2	2	3

At time i, n_i fishs are captured out of which m_i had already been previously captured. The total number of different fishes captured is thus equal to

$$\sum_{i=1}^{14} (n_i - m_i) = 138.$$

We consider here two type of models where

$$\begin{aligned} \pi(N, p_1, ..., p_I) &= \pi(N)\,\pi(p_1, ..., p_I) \\ &\propto \frac{1}{N}\mathbf{I}_{\mathbb{N}^*}(N) \prod_{k=1}^{I} \pi(p_k). \end{aligned}$$

In the first model, we assume

$$\pi(p_1, ..., p_I) = \prod_{k=1}^{I} \pi(p_k)$$

where

$$\pi(p_k) = \mathcal{B}e(\alpha, \beta)$$

with $(\alpha, \beta) = (1, 1)$ known.

In the second model, we assume that (α, β) are random and consider the following improper prior

$$\pi(\alpha, \beta) \propto (a + b)^{-2}\,\mathbf{I}(a, b > 1).$$

For these different priors, propose and write some Gibbs sampling algorithms to sample from $\pi(N, p_1, ..., p_I | n_1, ..., n_I)$ (when α, β are known) or $\pi(N, p_1, ..., p_I, \alpha, \beta | n_1, ..., n_I)$ (when α, β are unknown).

Show that it is possible to integrate out analytically $(p_1, ..., p_I)$ and propose and write alternative algorithms to approximate $\pi(N | n_1, ..., n_I)$ (when α, β are known) or $\pi(N, \alpha, \beta | n_1, ..., n_I)$ (when α, β are unknown).

Hand in any command scripts and programs you used, along with some informative output and plots, and your discussion.