Effect of bottom coverage to larval presence

Week2-ex3, solution

In this exercise you need the following special result. Assume we have a Binomial observation model

$$p(y|\theta, N) = \text{Bin}(y|N, \theta) \propto \theta^y (1-\theta)^{N-y}. \tag{1}$$

The number of trials, N, is considered to be fixed and the parameter θ is given a Beta prior

$$p(\theta) = \text{Beta}(\theta|\alpha, \beta) \propto \theta^{\alpha-1}(1-\theta)^{\beta-1}. \tag{2}$$

Comparing the observation model and prior gives insight that the Beta prior corresponds to $\alpha - 1$ prior successes and $\beta - 1$ prior failures. A special case is $\alpha = \beta = 1$ when the prior is uniform on the interval $[0, 1]$. The posterior distribution of θ is now also Beta distribution (see BDA 3, Chapter 2 for more details)

$$\theta|N, y \sim \text{Beta}(\alpha + y, \beta + N - y). \tag{3}$$

Note! The Beta posterior arizes only if we use Beta prior. If we use some other prior than Beta distribution the posterior will not be Beta distribution.

Using the above result solve the below problem.

Problem statement

We analyze the data presented by Veneranta et al. (2013). White fish is a fresh water origin fish species that is found also from the northern parts of the brackish water Gulf of Bothnia. The species is important for both commercial and recreational fisheries. White fish spawn in shallow coastal areas and former studies suggests that the survival of white fish larvae is decreased by algal or other bottom vegetation which have been increasing throughout Finnish and Swedish coastal region due to eutrophication. Hence, we want to study whether there is difference in the probability of presence of white fish in clear and vegetated areas.

A number of sites (n=653) along the Finnish and Swedish coastal region in the Gulf of Bothnia were sampled during 2009-2011. At each site, researchers sampled a volume of water using a fine meshed net and recorded whether or not white fish larvae were detected or not. Each site was classified with COVERAGE: 0 = clear and 1 = covered with vegetation. The data contains also other covariates and spatial information but these will be considered in later weeks.

The data (contingency table) for this exercise can be downloaded and formed as follows

```
data = read.csv("white_fishes_data.csv")

y = table(data$WHIBIN, data$BOTTOMCOV)
colnames(y) <- c("COV=0","COV=1")
rownames(y) <- c("y=0","y=1")
print(y)
```

```
##
##       COV=0 COV=1
##   y=0    65   104
##   y=1   212   121
```

The variable y groups the sampling sites into locations with respect to the vegetation cover (COV=0 vs. COV=1) and occurrence of white fish larvae (y=0 vs. y=1).

Let's assume that the outcomes of sampling occasions (presence/absence of whitefish) are independent Bernoulli (Binomial with sample size 1) distributed random variables with success probabilities θ_0 for sites with no vegetation cover and θ_1 for sites with vegetation cover.
Let's further assume that there is no prior information on θ_c, $c \in \{0, 1\}$ so that their prior is uniform between 0 and 1, and that the parameters θ_c are mutually independent.

1. Write down the equation for the posterior distribution for both θ_0 and θ_1
2. Sample random draws from both posterior distributions, draw a histogram of the samples and report the posterior mean and standard deviation.
3. Visualize the posterior distribution of $\phi = \theta_0 - \theta_1$ and calculate the posterior probability that $\theta_1 < \theta_0$
4. Analyze and discuss the sensitivity of the results to the choice of the prior distribution

Grading

Total 10 points Three points from both part 1 and 4. Two points from both part 2 and 3. In all sub-questions you should give point if the idea of the solution is correct. One extra point if the solution is in principle correct but contains minor typo/bug. Full points from totally correct answer.

References

Lari Veneranta, Richard Hudd and Jarno Vanhatalo (2013). Reproduction areas of sea-spawning Coregonids reflect the environment in shallow coastal waters. Marine Ecology Progress Series, 477:231-250. http://www.int-res.com/abstracts/meps/v477/p231-250/