Bayesian Sample Size Calculation

March 14, 2019

Background

As part of a rockfish research initiative we want to estimate the length and age at which there is a 50% probability a rockfish is mature at locations in and around PWS. Measurements will be binary mature/not mature, so we will use a logistic regression to model the probability of being mature as a function of age and length. We will be using a bayesian approach to estimate the L50 and A50 parameters.

The Problem

We want to determine how many fish must be sampled in order to obtain a meaningfully precise credible interval.

The Solution

First we will note that Oregon FWS (Hannah 2009) estimated rockfish length and age at maturity (A50 and L50) by performing a logistic regression on 148 Yelloweye Rockfish.

For $P(mature) = \frac{1}{1+exp(-(\beta_0+\beta 1*length))}$ they obtained estimates of $\beta_0 = -16.948$ and $\beta_1 = .437$

Using this information we will perform the following steps to determine the precision of estimates for a given sample sizes.

1. Simulate rockfish lengths, the number of which would be your prospective sample size.
2. Using the estimates from Hannah 2009, simulate maturity status (0 or 1) from a bernoulli function with $P(1) = \frac{1}{1+exp(-(-16.948+.437*length))}$
3. Using the above simulated maturity data, obtain a credible interval for L50, recording the width of the interval as a percentage of the estimate.
4. Repeat steps 1-3 1000 times and use the maximum width obtained out of all of the step threes as an upper bound on the possible width of the credible interval.

Then we can say our objective is "To estimate L50 (A50) such that our estimate is within X% of the true value with probability P" and have a reasonable estimate of the sample size required to meet such a criteria.

Example:

Suppose we want to know how precise our estimate will be with 100 measurements.

1. First we simulate 100 lengths
2. Then simulate 100 0/1's using probability function obtained by Hannah 2009
3. Use the above simulated maturity data to obtain a credible interval for L50, recording the width
4. Repeat 1000 times and use the maximum observed interval width, in terms of percent of our estimate, as our upper bound (The X% above).