Stat 343: Introduction to Maximum Likelihood Estimation; Capture/Recapture

Evan Ray

January 23, 2018

Introduction

Often, ecologists want to estimate the total number (population size) of a certain kind of animal in a given area. If the population size remains constant throughout the sampling period, the capture-recapture method is often used. In this method, one samples c animals and releases them after "tagging" them. After sufficient time has passed for the tagged animals to thoroughly mix with the untagged population, n animals are captured, the number of tagged animals in the n units is counted, and based on this information the unknown population size N is estimated.

In this lab, we will explore the capture/recapture method in the context of estimating the total number of goldfish in a pond.

- Research Question: How big is the population?
- Population parameter to estimate: N, the population size

1. Set up

Pour 100 Cheddar goldfish into a ziplock bag. This represents the population of fish in the pond. For the rest of this exercise, we're going to pretend that we don't know the total number of fish in the pond and we're going to try to estimate that number.

2. Capture (or Mark)

(a) Use a dixie cup to draw a random sample of fish from the pond.

These are the "captured" fish. We're going to mark each of these fish with an identifying tag so that we will know we've captured this fish before if we see it again in the next step. If these were real fish, we would use physical tags, but in this simulation we'll mark the fish by making them change color.

(b) Count the number of fish you captured in this first sample: $c =$ ____________

(c) Replace each of the fish you "captured" with a pretzel fish.

The new goldfish replace the old ones – don't put the old ones back into the pond!

3. Recapture

Stir the fish around in the pond so that the marked and unmarked fish are well mixed together. You're about to draw a new random sample of fish from the lake and count the number of fish in the second step that are marked. Before you do that, let's think about what we're doing from a probabilistic modeling point of view.

(a) Let the random variable X represent the number of tagged fish in a sample of size n from a population of size N where c of the fish in the population are marked. What distribution does X follow?

$X \sim$

OK, now go ahead and take a second sample, and record the following information about this second sample:

(a) Count the total number of fish you captured in this second sample: $n =$ ____________

(b) Count the number of fish in this second sample that are marked (pretzel): $x =$ ____________

4. Estimate the proportion of fish in the pond that were captured in the Capture step

Our goal in all of this was to use the data collected in the capture/recapture process to estimate the total population size.

Here's what we have so far:

- N: population size; this is the "unknown" quantity we want to estimate.
- c: total number of fish that were marked in step 2.
- n: sample size for the recapture step, from step 3 (a).
- x: number of fish that were marked in the recapture step, from 3 (b).

Intuitively, what seems like a good estimator for N? Write this out using the symbols c, n, and x. Then, plug in the numbers for those quantities that you got above to calculate your estimate.

5. A second approach to estimation...

Let's go back in time to just before the recapture step. So, you've just captured c fish, marked them, and released them back into the pond.

(a) Let the random variable X represent the number of tagged fish in a sample of size n from a population of size N. What distribution does X follow?

(b) What is the probability mass function of X? Write this in terms of the symbols N, n, x, and c.

$P(X = x) =$

(c) For simplicity, let's assume momentarily that in a sample of three fish, we catch one of four total tagged fish. That is, $n = 3$, $x = 1$, and $c = 4$. We want to estimate the total number of fish in this particular population, N. Given that we observed one tagged fish in a sample of size $n = 3$, what population size, N, would make this observation most likely? How could we find out?