Topic 13
Method of Moments
Examples

Outline

Mark and Recapture

The Lincoln-Peterson method of mark and recapture

- The size of an animal population in a habitat of interest is an important question in conservation biology.
- In many case, Individuals are often too difficult to find, a census is not feasible.
- One estimation technique is to capture some of the animals, mark them and release them back into the wild to mix randomly with the population.
- Some time later, a second capture from the population is made.

Mark and Recapture

Some of the animals were not in the first capture and some, which are tagged, are recaptured. Let

- t be the number captured and tagged,
- k be the number in the second capture,
- r be the number in the second capture that are tagged, and let
- N be the total population size.

Thus, t and k is under the control of the experimenter. The value of r is random and the populations size N is the parameter to be estimated.

Mark and Recapture

We can guess the the estimate of N by considering two proportions.

the proportion of the tagged fish in the second capture	$\approx$	the proportion of tagged fish in the population

$$\frac{r}{k} \approx \frac{t}{N}$$

This can be solved for N to find

$$N \approx \frac{kt}{r}.$$

The advantage of obtaining this as a method of moments estimator is that we evaluate the precision of this estimator by determining, for example, its variance.

Mark and Recapture

To begin, let

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th individual in the second capture has a tag.} \\ 0 & \text{if the } i\text{-th individual in the second capture does not have a tag.} \end{cases}$$

The X_i are Bernoulli random variables with success probability $P\{X_i = 1\} = t/N$. The number of tagged individuals is $X = X_1 + X_2 + \cdots + X_k$ and the expected number of tagged individuals is

$$\mu = EX = EX_1 + EX_2 + \cdots + EX_k = \frac{t}{N} + \frac{t}{N} + \cdots + \frac{t}{N} = \frac{kt}{N}.$$

The proportion of tagged individuals, $\bar{X} = (X_1 + \cdots + X_k)/k$, has expected value

$$E\bar{X} = \frac{\mu}{k} = \frac{t}{N}. \quad \text{Thus, } N = \frac{kt}{\mu}.$$

Mark and Recapture

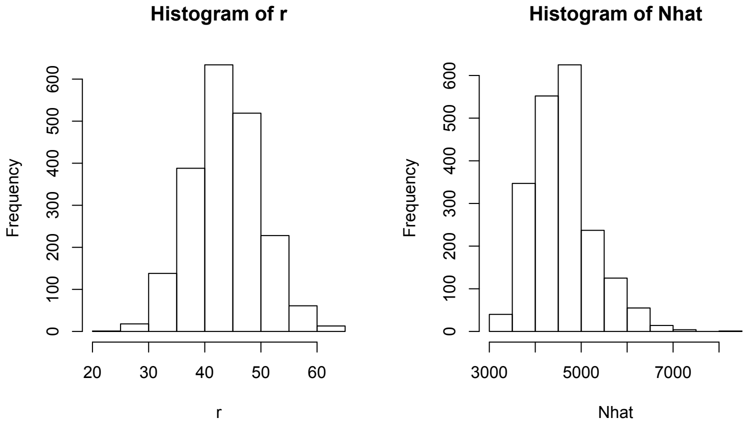
Now in this case, we are estimating μ, the mean number recaptured with r, the actual number recaptured. So, to obtain the estimate $\hat{N}$. we replace μ with the previous equation by r.

$$\hat{N} = \frac{kt}{r}.$$

We simulate the process in a lake having 4500 fish.

```
> N<-4500;t<-400;k<-500      #population 4500, 400 tagged, recapture 500
> r<-rep(0,2000)             #set a vector of zeros for 2000 simulations
> fish<-c(rep(1,t),rep(0,N-t)) #tag t fish
> for (j in 1:2000){r[j]<-sum(sample(fish,k))}
> Nhat<-k*t/r                #compute estimate of population
> mean(Nhat);sd(Nhat)
[1] 4606.933
[1] 666.1918
```

Mark and Recapture



Exercise. Describe the histograms above. Comment of the mean and standard deviation of the estimate $\hat{N}$.

Distribution of Fitness Effects

- Fitness is a central concept in the theory of evolution.
- Relative fitness is quantified as the average number of surviving progeny of a particular genotype compared with average number of surviving progeny of competing genotypes after a single generation.
- Consequently, the distribution of fitness effects for newly arising mutations is a basic question in evolution.

Even though more fit genes are more likely to fix in a populations, humans have, just by chance, genes that are deleterious. We will consider, in humans, the distribution of fitness effects of such mutations

Distribution of Fitness Effects

Our model for this distribution will be the gamma family of random variables.

A $\Gamma(\alpha, \beta)$ random variable has mean α/β and variance α/β^2. Because we have two parameters, the method of moments methodology requires us, in step 1, to determine the first two moments.

$$\mu_1 = E_{(\alpha,\beta)}X_1 = \frac{\alpha}{\beta}$$

$$\begin{aligned} \mu_2 = E_{(\alpha,\beta)}X_1^2 &= \mathrm{Var}_{(\alpha,\beta)}(X_1) + (E_{(\alpha,\beta)}X_1)^2 \\ &= \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 = \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} = \frac{\alpha(1+\alpha)}{\beta^2}. \end{aligned}$$

NB. $\mathrm{Var}(Y) = EY^2 - (EY)^2$. So, $EY^2 = \mathrm{Var}(Y) + (EY)^2$.

Distribution of Fitness Effects

The first two moments

$$\mu_1 = \frac{\alpha}{\beta} \quad \text{and} \quad \mu_2 = \frac{\alpha}{\beta^2} + \frac{\alpha^2}{\beta^2} = \frac{\alpha(1+\alpha)}{\beta^2}$$

For step 2, we solve for α and β. Note that

$$\mu_2 - \mu_1^2 = \frac{\alpha}{\beta^2},$$

$$\frac{\mu_1}{\mu_2 - \mu_1^2} = \frac{\alpha/\beta}{\alpha/\beta^2} = \beta,$$

and

$$\mu_1 \cdot \frac{\mu_1}{\mu_2 - \mu_1^2} = \frac{\alpha}{\beta} \cdot \beta = \alpha, \quad \text{or} \quad \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

Distribution of Fitness Effects

$$\beta = \frac{\mu_1}{\mu_2 - \mu_1^2} \quad \text{and} \quad \alpha = \beta\mu_1 = \frac{\mu_1^2}{\mu_2 - \mu_1^2}.$$

For step 3, set the first two sample moments

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{and} \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$$

to obtain estimates

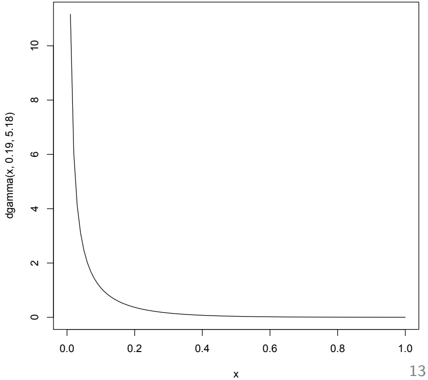
$$\hat{\beta} = \frac{\bar{x}}{\overline{x^2} - (\bar{x})^2} \quad \text{and} \quad \hat{\alpha} = \hat{\beta}\bar{x} = \frac{(\bar{x})^2}{\overline{x^2} - (\bar{x})^2}$$

as required in step 4.

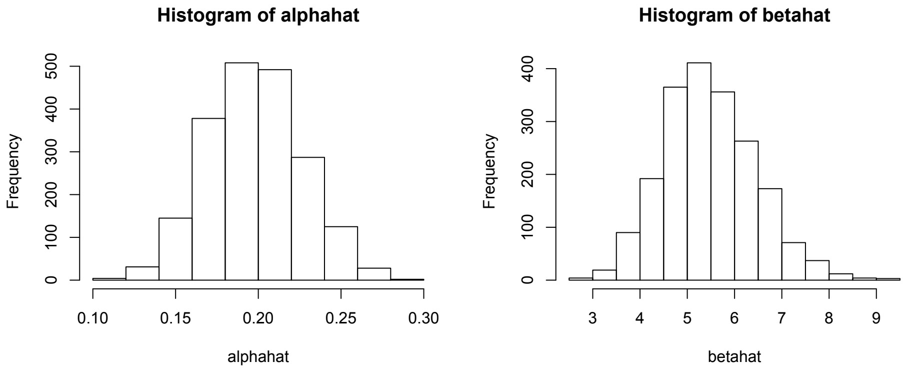
Distribution of Fitness Effects

In a recent manuscript based on the exomes of 44 unrelated Yoruban females, we found estimates of $\hat{\alpha} = 0.19$ and $\hat{\beta} = 5.18$. To investigate the method of moments estimates on simulated data using R, we consider 2000 repetitions of 500 independent observations of a $\Gamma(0.19, 5.18)$ random variable.

```
> xbar <- rep(0,2000); x2bar <- rep(0,2000)
> for (i in 1:2000){x<-rgamma(500,0.19,5.18);
  xbar[i]<-mean(x);x2bar[i]<-mean(x^2)}
> betahat <- xbar/(x2bar-(xbar)^2)
> alphahat <- betahat*xbar
> mean(alphahat);sd(alphahat)
[1] 0.1979031
[1] 0.02837945
> mean(betahat);sd(betahat)
[1] 5.472508
[1] 0.9769274
```



Distribution of Fitness Effects



Exercise. Describe the histograms above. Comment of the mean and standard deviation of the estimates $\hat{\alpha}$ and $\hat{\beta}$. The correlation $r(\hat{\alpha}, \hat{\beta}) = 0.810$ for this simulation. What does this tells us about the estimates? Repeat the simulation and give a plot of $\hat{\alpha}$ versus $\hat{\beta}$.