Chapter 7
Point Estimation
Method of Moments

Outline

Parameter Estimation

For parameter estimation, we consider $X = (X_1, \ldots, X_n)$, independent random variables chosen according to one of a family of probabilities P_θ where θ is element from the parameter space Θ. Based on our analysis, we choose an estimator $\hat{\theta}(X)$. If the data $\mathbf{x}$ takes on the values $x_1, x_2, \ldots, x_n$, then

$$\hat{\theta}(x_1, x_2, \ldots, x_n)$$

is called the estimate of θ. Thus we have three closely related objects.

1. θ - the parameter, an element of the parameter space, is a number or a vector.
2. $\hat{\theta}(x_1, x_2, \ldots, x_n)$ - the estimate, is a number or a vector obtained by evaluating the estimator on the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.
3. $\hat{\theta}(X_1, \ldots, X_n)$ - the estimator, is a random variable. We will analyze the distribution of this random variable to decide how well it performs in estimating θ.

Classical Statistics

In classical statistics, the state of nature is assumed to be fixed, but unknown to us. Thus, one goal of estimation is to determine which of the P_θ is the source of the data. The estimate is a statistic

$$\hat{\theta} : \text{data} \to \Theta.$$

For estimation procedures, the classical approach to statistics is based on two fundamental questions:

- How do we determine estimators?
- How do we evaluate estimators?
  - Does this estimator in any way systematically under or over estimate the parameter?
  - Does it has large or small variance?
  - How does it compare to a notion of best possible estimator?
  - How easy is it to determine and to compute?
  - How does the procedure improve with increased sample size?

Method of Moments

Method of moments estimation is based solely on the law of large numbers,

Let $M_1, M_2, \ldots$ be independent random variables having a common distribution possessing a mean μ_M. Then the sample means converge to the distributional mean as the number of observations increase.

$$\bar{M}_n = \frac{1}{n}\sum_{i=1}^{n} M_i \to \mu_M \quad \text{as } n \to \infty$$

almost surely and in mean.

In addition, if the random variables in this sequence *fail* to have a mean, then the limit will *fail* to exist.

Procedure

- Step 1. If the model is based on a parametric family of densities $f_X(x|\theta)$ with a d-dimensional parameter space $(\theta_1, \theta_2 \ldots, \theta_d)$, we compute

$$\mu_m = EX^m = k_m(\theta) = \int_S x^m f_X(x|\theta)\ \nu(dx), \quad m = 1, \ldots, d$$

the first d moments,

$$\mu_1 = k_1(\theta_1, \theta_2 \ldots, \theta_d), \quad \mu_2 = k_2(\theta_1, \theta_2 \ldots, \theta_d), \quad \ldots, \quad \mu_d = k_d(\theta_1, \theta_2 \ldots, \theta_d),$$

obtaining d equations in d unknowns.

Example

Let $X_1, X_2, \ldots, X_n$ be a simple random sample of Pareto random variables with density

$$f_X(x|\beta) = \frac{\beta}{x^{\beta+1}}, \quad x > 1, \qquad \beta > 0.$$

The cumulative distribution function is

$$F_X(x) = 1 - x^{-\beta}, \quad x > 1.$$

The mean and the variance are, respectively,

$$\mu = \frac{\beta}{\beta - 1}, \qquad \sigma^2 = \frac{\beta}{(\beta-1)^2(\beta-2)}.$$

In this situation, we have one parameter, namely β. Thus, in step 1, we will only need to determine the first moment

$$\mu_1 = \mu = k_1(\beta) = \frac{\beta}{\beta - 1}$$

to find the method of moments estimator $\hat{\beta}$ for β.

Procedure

- Step 2. We then solve for the d parameters as a function of the moments.

$$\theta_1 = g_1(\mu_1, \mu_2, \cdots, \mu_d), \quad \theta_2 = g_2(\mu_1, \mu_2, \cdots, \mu_d),$$
$$\ldots, \quad \theta_d = g_d(\mu_1, \mu_2, \cdots, \mu_d).$$

- Step 3. Now, based on the data $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, we compute the first d sample moments,

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad \overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2, \quad \ldots, \quad \overline{x^d} = \frac{1}{n}\sum_{i=1}^{n} x_i^d.$$

Example

Exercise. If

$$\mu = \frac{\beta}{\beta - 1}, \text{ show that } \beta = \frac{\mu}{\mu - 1}.$$

For step 2, we solve for β as a function of the mean μ.

$$\beta = g_1(\mu) = \frac{\mu}{\mu - 1}.$$

For step 3, we compute the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

Procedure

- Step 4. Using the law of large numbers, we have, for each moment, $m = 1, \ldots, d$, that

$$\mu_m \approx \overline{x^m}.$$

For the equations derived in step 2, we replace the distributional moments μ_m by the sample moments $\overline{x^m}$ to give us formulas for the method of moment estimates

$$(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_d).$$

For the data $\mathbf{x}$, these estimates are

$$\hat{\theta}_1(\mathbf{x}) = g_1(\bar{x}, \overline{x^2}, \cdots, \overline{x^d}), \quad \hat{\theta}_2(\mathbf{x}) = g_2(\bar{x}, \overline{x^2}, \cdots, \overline{x^d}),$$
$$\ldots, \quad \hat{\theta}_d(\mathbf{x}) = g_d(\bar{x}, \overline{x^2}, \cdots, \overline{x^d}).$$

Example

Consequently, a method of moments estimate for β is obtained by replacing the distributional mean μ by the sample mean $\bar{x}$ in the equation for $g_1(\mu)$. Thus,

$$\hat{\beta} = \frac{\bar{x}}{\bar{x} - 1}.$$

A good estimator should have a small variance. To use the delta method to estimate the variance of $\hat{\beta}$,

$$\sigma^2_{\hat{\beta}} \approx g_1'(\mu)^2 \frac{\sigma^2}{n}.$$

We compute

$$g_1(\mu) = \frac{\mu}{\mu - 1} \quad \text{and so} \quad g_1'(\mu) = -\frac{1}{(\mu - 1)^2}$$

giving

$$g_1'\left(\frac{\beta}{\beta - 1}\right) = -\frac{1}{(\frac{\beta}{\beta-1} - 1)^2} = -\frac{(\beta - 1)^2}{(\beta - (\beta - 1))^2} = -(\beta - 1)^2.$$

Example

We find that $\hat{\beta}$ has mean approximately equal to β and variance

$$\sigma^2_{\hat{\beta}} \approx g_1'(\mu)^2 \frac{\sigma^2}{n} = (\beta-1)^4 \frac{\beta}{n(\beta-1)^2(\beta-2)} = \frac{\beta(\beta-1)^2}{n(\beta-2)}.$$

Let's consider the case with $\beta = 4$ and $n = 225$. Then,

$$\sigma^2_{\hat{\beta}} \approx \frac{4\cdot 3^2}{225\cdot 2} = \frac{36}{450} = \frac{2}{25}, \quad \sigma_{\hat{\beta}} \approx \frac{\sqrt{2}}{5} = 0.283.$$

To simulate, we use the probability transform

$$u = F_X(x) = 1 - x^{-\beta}, \quad \text{then} \quad x = (1-u)^{-1/\beta} = 1/\sqrt[\beta]{(1-u)}.$$

Note that if U_i are $U(0,1)$ random variables, then $1/\sqrt[\beta]{(1-U_1)}, 1/\sqrt[\beta]{(1-U_2)}, \cdots$ have the appropriate Pareto distribution.

Example

```
> paretobar<-rep(0,1000)
> for (i in 1:1000){u<-runif(225);
   pareto<-1/(1-u)^(1/4);
   paretobar[i]<-mean(pareto)}
> betahat<-paretobar/(paretobar-1)
> mean(betahat)
[1] 4.03508
> sd(betahat)
[1] 0.2833142
```

Note that the mean is above 4, but the standard deviation is very close to the value given by the delta method.

Exercise. Reproduce the simulation above and compare. Simulate using a different value for β.

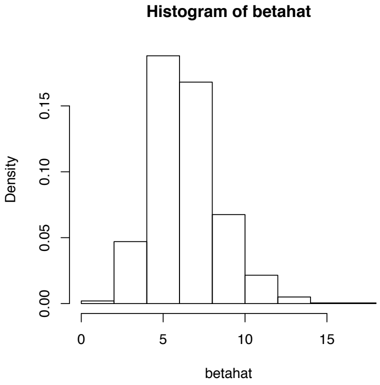


Figure: 1000 simulations for the method of moments estimate for the case $\beta = 4$.

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 and 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.

$$\begin{array}{rcl} \text{the proportion of the tagged fish} & \approx & \text{the proportion of tagged fish} \\ \text{in the second capture} & & \text{in the population} \end{array}$$

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

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

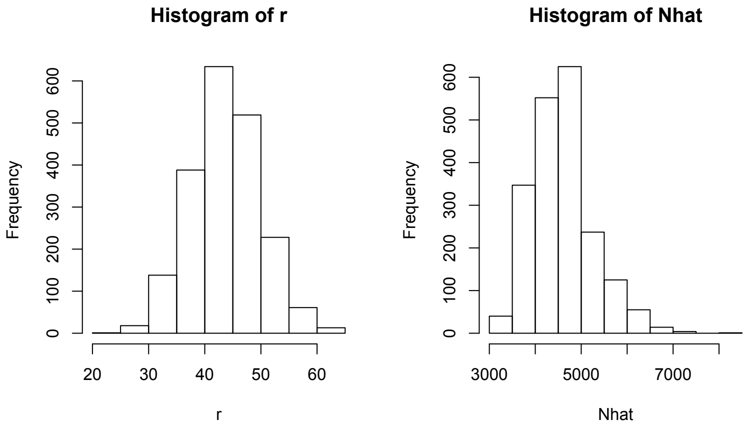
Now in this case, we are estimating μ, the mean number recaptured with r, the actual number recaptured. So, to obtain the estimate $\hat{N}$. we replace μ with the previous equation by r.

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

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 $\hat{N}$.

Monsoon Rains

Monsoon is used to describe the rainy phase for seasonal changes in atmospheric circulation and precipitation associated with the asymmetric heating of land and sea. 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, $EY^2 = \mathrm{Var}(Y) + (EY)^2$.

Monsoon Rains

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

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

Monsoon Rains

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

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.

Monsoon Rains

In 2017, Tucson, Arizona had 17 summer monsoon rainstorms. The data are the rainfall in millimeters.

```
x<-c(3,15,1,37,5,1,8,11,6,9,12,
   35,22,3,38,1,2)
> (xbar<-mean(x));(s2<-var(x))
[1] 12.29412
[1] 167.0956
> (betahat<-xbar/s2);
  (alphahat<-betahat*xbar)
[1] 0.07357536
[1] 0.9045441
```

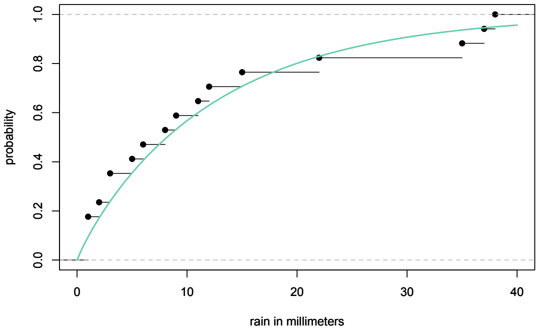


Figure: Comparison of the empirical distribution function for the monsoon rainfall data and the plot of the estimated gamma distribution function.