Chapter 8 Estimation

We want to make inference about a population based on information contained in a random sample.

A random sample $Y_1 \ldots Y_n$ consists of iid r.v's with a distribution $f(y)$. Usually, we assume that the distribution $f(y)$ is determined by unknown quantities called parameters.

We denote θ be the unknown parameter determining the distribution of the population

Let Θ be the set of all values of the parameter θ.

Θ is called the parameter space.

We assume that the distribution of the data $Y_1 \ldots Y_n$ is $f(y, \theta)$ for some unknown value $\theta \in \Theta$.

An estimator $\hat{\theta}$ is a rule that tells how to calculate the value of an estimate. $\hat{\theta}$ is a function of the data $Y_1 \ldots Y_n$ and known parameters. An estimate is the value of the estimator obtained from the sample. An estimate is the value of the function determined by the estimator when applied to a particular data

Example

We want to estimate Y be the number of mileages of a model of a car before a major repair is made.

To being able to use the theory we know, we assume that Y has a normal distribution distribution with unknown mean μ and unknown variance σ^2.

The parameter is $\theta = (\mu, \sigma^2)$

The parameter space is $\Theta = \{ (\mu, \sigma^2) : \mu \in \mathbb{R}, \sigma^2 > 0 \}$

An estimator of μ is $\bar{Y} = \frac{1}{n} \sum_{j=1}^{n} Y_j$.

$\bar{Y}$ is called the sample mean,

μ is called the population mean.

If the value we observe are

30,000, 40,000, 40,000, 30,000, 35,000

then an estimate of μ is

$$\bar{y} = \frac{1}{5}(30{,}000 + 40{,}000 + 40{,}000 + 30{,}000 + 35{,}000) = 35000$$

$$f(y, \theta) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}}, \quad y \in \mathbb{R}, \quad \theta = (\mu, \sigma^2)$$

Example

In the primary Democratic Elections, there are two candidates Kerry and Edwards. We want to determine p, the probability that a voter will vote for Kerry. To do that we sample n voters. Let $Y_j = \begin{cases} 1 & \text{if voter } j \text{ prefers Kerry} \\ 0 & \text{if not} \end{cases}$

Then, $\hat{p} = \frac{1}{n} \sum_{j=1}^{n} Y_j = \frac{Y}{n}$ is an estimator of p.

Here, we assume that $Y_1, \dots, Y_n$ are iid r.v and Y has a binomial distribution with parameters n and p

$$f_Y(y|p) = \binom{n}{y} p^y (1-p)^{1-y}, \text{ for } y = 0, 1, \dots n.$$

Assessing point estimators

A point estimator $\hat{\theta}$ is an estimator, with the same dimension as θ.

$\hat{\theta}(Y_1 \ldots Y_n)$ is a guess on the unknown value θ.

There are many different properties we can consider to study how good an estimator is.

Unbiased estimators

$\hat{\theta}$ is an unbiased estimator of θ if $E_\theta[\hat{\theta}] = \theta$

$\hat{\theta}$ is a biased estimator of θ, if $\hat{\theta}$ is not unbiased.

The bias B of a point estimator $\hat{\theta}$ is given by

$B = E[\hat{\theta}] - \theta$

Unbiased estimators are preferred

Mean Square error

The error of estimation ε is the distance between an estimator and its target parameter. That is $\varepsilon = |\hat{\theta} - \theta|$.

The mean square error of a point estimator $\hat{\theta}$ is the expected value of $(\hat{\theta} - \theta)^2$

$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$

Since, $Var(Y) = E[Y^2] - (E[Y])^2$, $E[Y^2] = Var(Y) + (E[Y])^2$

and $MSE(\hat{\theta}) = E[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}-\theta) + (E[\hat{\theta}-\theta])^2 = Var(\hat{\theta}) + (B(\theta))^2$

Estimators with small MSE are preferred.

<u>Goal</u> Given a parametric family $\{f(y|\theta) : \theta \in \Theta\}$ find the unbiased estimator, whose MSE is the minimum over all possible unbiased estimators.