10 Two additional point estimation method

10.1 Maximum Likelihood Estimators

Method of Maximum Likelihood (ML) produces an estimate of the population parameter θ which corresponds to the density function of a random sample statistic. Another method of point estimation through what is called the 'likelihood function'

  ▷ The likelihood function is identical in functional form to the joint density function of the random sample.

Joint Density Function: is nothing but a function of random variables $(x_1, x_2, \ldots x$ given values of the parameters $(\theta_1, \theta_2, \ldots)$ Example: What are the possible values of incomes of a person in India (x_1)

  ▷ We know that the average income per annum is $\theta_1 = 16000 Rs$

  ▷ We also know that the standard deviation is $\theta_2 = 5000 Rs$

The likelihood function is similar in functional form to the joint density function of the random sample. Likelihood function: is just the reverse of JDF - it is a function of the values of the parameters $(\theta_1, \theta_2, \ldots)$, given values of the random variable outcomes $(x_1, x_2, \ldots x_n)$. Denoted as $L(\theta; x_i)$ Estimation: The maximum likelihood estimate of θ is the solution of the maximization problem

$$argmax_{\Theta \in \Omega}\{L(\theta; x_i)\},$$

where Ω is the appropriate parameter space. Estimate: The maximum likelihood estimate is thus defined as $\hat{\theta} = \theta(x_i)$

▷ Idea: is that we need to find parameters which increase the likelihood of our sample being drawn from an identifiable population, by maximizing the likelihood function

○ How do we obtain the likelihood function?

▷ Exponential Distribution Example:

○ Let $X \in (x_1, \ldots, x_n)$ be a random sample from an exponential population distribution representing the operating time until a work stoppage occurs on an assembly line,

$$x_i \sim \frac{exp(-x_i/\theta)}{\theta}$$

where $\hat{\theta}$ is the MLE for θ which is the mean operating time untill work gets stopped in the line.

▷ The Likelihood function is given by:

$$L(\theta; x_i) = \left(\frac{1}{\theta} e^{\frac{-x_1}{\theta}} \right) \left(\frac{1}{\theta} e^{\frac{-x_2}{\theta}} \right) \cdots \left(\frac{1}{\theta} e^{\frac{-x_n}{\theta}} \right)$$

$$= \left(\frac{1}{\theta^n} e^{\frac{\sum_{i=1}^{n} -x_i}{\theta}} \right)$$

▷ Calculations are often simplified by taking log of the likelihood function and then maximizing

$$lnL(\theta; x_i) = -n \ln \theta + \frac{\sum_{i=1}^{n} -x_i}{\theta}$$

▷ Estimation implies maximizing this function and setting it equal to zero to obtain $\hat{\theta}$

$$\frac{\partial lnL(\theta; x_i)}{\partial \theta} = \frac{-n}{\theta} - \frac{\sum_{i=1}^{n} -x_i}{\theta^2} = 0$$

▷ Simplifying this gives us the maximum likelihood estimate of θ

$$\hat{\theta} = \frac{\sum_{i=1}^{n} x_i}{n}$$

▷ Example: Suppose the weights of randomly selected female college students are normally distributed with unknown mean μ and standard deviation σ.

▷ A random sample of 10 female college students yielded the following weights (in pounds)

$$
\begin{array}{ccccc}
115 & 122 & 130 & 127 & 149 \\
160 & 152 & 138 & 149 & 180
\end{array}
$$

▷ Identify the likelihood function and find the maximum likelihood estimator of μ, the mean weight of all female college students.

▷ Recall: PDF of the normal distribution

$$f(x_i; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} exp\left[- \frac{(x_i - \mu)^2}{2\sigma^2} \right] \qquad x \in (-\infty, \infty)$$

▷ Likelihood function is given by:

$$L(\mu, \sigma; x_i) = \left(\frac{1}{\sigma\sqrt{2\pi}} exp\left[- \frac{(x_1 - \mu)^2}{2\sigma^2} \right] \right)$$
$$\left(\frac{1}{\sigma\sqrt{2\pi}} exp\left[- \frac{(x_2 - \mu)^2}{2\sigma^2} \right] \right)$$
$$\vdots$$
$$\frac{1}{\sigma\sqrt{2\pi}} exp\left[- \frac{(x_n - \mu)^2}{2\sigma^2} \right]$$

▷ Likelihood function is simplified to:

$$L(\mu, \sigma; x_i) = \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} exp\left[- \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2} \right]$$

▷ Again, take the log:

$$\ln L(\mu, \sigma; x_i) = -n \ln \sigma - \frac{n}{2} \ln 2\pi - \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{2\sigma^2}$$

▷ Maximization by choosing two parameters this time i.e. μ and σ, we have two first order conditions

$$\frac{\partial \ln L}{\partial \mu} = \frac{\sum_{i=1}^{n}(x_i - \mu)}{\sigma^2} = 0$$

$$\frac{\partial \ln L}{\partial \sigma} = \frac{-n}{\sigma} + \frac{\sum_{i=1}^{n}(x_i - \mu)^2}{\sigma^3} = 0$$

▷ Simplifying for μ to obtain the MLE estimate from first FOC:

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n}$$

▷ Simplifying for σ to obtain the MLE estimate from second FOC:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

▷ μ MLE estimate for our weight problem

$$\hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = 142.2$$

▷ σ MLE estimate for our weight problem

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n} = 347.96$$

  ○ Why is the variance so large?

10.2 Method of Moments Estimators

▷ Method of moments involves equating sample moments with theoretical moments

    ○ Originally suggested long ago by Karl Pearson, is based on iid sampling and moment conditions that are equal in number to the number of unknown parameters.

▷ The moment conditions take the general form

$$E(g(x_i, \theta)) = 0$$

where it is understood that the expectation is taken with respect to $f(x_i; \theta)$ (the pdf).

▷ Generalized Method of Moments: is an estimation procedure which does not require iid assumption.

▷ Theoretical moments (recall)

- $E(x^k)$ is the kth theoretical moment of the distribution about origin

- $E[(x_i - \mu)^k]$ is the kth theoretical moment of the distribution about mean

▷ Sample counterpart for theoretical moments:

- Sample counterpart about origin is $M_k = \frac{1}{n} \sum_{i=1}^{n} x_i^k$ is the kth sample moment

- Sample counterpart about mean is $M_k = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^k$ is the kth sample moment

▷ **Theoretical Example**: Let $(x_1, x_2, \ldots, x_n)$ be a random sample from an unknown population which follows a normal distribution

▷ **Objective**: desire to estimate the parameters (μ, σ^2) i.e. how much is the average, and how much is the expected dispersion.

▷ **Step 1**: Obtain the first theoretical moment about origin for a normal distribution:

$$E(X) = \mu$$

▷ Step 2: Obtain the second theoretical moment about origin:

$$E(X^2) = \sigma^2 + \mu^2$$

▷ Step 3: Equate the first sample moment to its theoretical moment to obtain an estimate of population average as:

$$E(X) = \mu = \frac{1}{n} \sum_{i=1}^{n} x_i$$

$$E(X) = \bar{x}$$

Thus the moment estimator for population mean is the sample mean.

▷ Step 4: Equate the second sample moment to its theoretical moment to obtain an estimate of population variance as:

$$E(X^2) = \sigma^2 + \mu^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2$$

▷ But we know that $\mu = \bar{x}$:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \mu^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i^2 - \bar{x}^2$$

$$= \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

last equality because $2 x_i \bar{x} = 0$ when moments are near origin (use the formula $(a - b)^2$).

▷ Summarizing: Moments about the mean: for an estimate of the mean, equate the first sample moment about origin to the first theoretical moment

$$m_1 = E(x)$$

$$\frac{1}{n} \sum_{i=1}^{n} x_i = \mu$$

▷ For an estimate of the second central moment: equate the second sample moment about sample mean with corresponding population moment

$$m_2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 = E[(x_i - \mu)^2]$$

▷ Exponential Distribution: suppose that $(x_1, \ldots, x_n)$ is a random sample from an exponential population so that their PDF is

$$f(x|\theta) = \theta e^{-\theta x} \qquad x > 0$$

▷ Let us construct a method of moments estimator for the unobserved mean of this distribution. First theoretical moment:

$$m_1 = E(x) = \theta \int_0^\infty x e^{-\theta x} dx = \frac{1}{\theta}$$

▷ Its sample counterpart is:

$$\hat{m}_1 = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

▷ Setting both equal to determine method of moments estimator

$$m_1 = \hat{m}_1$$

implies that

$$\frac{1}{\theta} = \bar{x}$$

▷ Estimator for the average implies:

$$\hat{\theta} = \frac{1}{\bar{x}} = n\left(\sum_{i=1}^{n} x_i\right)^{-1}$$