Chapter 5

Sampling Distributions and Point Estimation of Parameters

CLO5	Define important properties of point estimators and construct point estimators using maximum likelihood.

1. Introduction

- We learned that a probability distribution provides a way to assign probabilities to the possible values of the random variable.
- In this chapter, we discuss about probability distributions where statistics, such as the mean, will be the random variable.
- We use probability distributions to make statements regarding the statistic.

2. Sampling distributions and the central limit theorem

- **Random sample:** The random variables $X_1, X_2, \dots, X_n$ are a random sample of size *n* if they are *independent* random variables, and every X_i has the *same probability distribution* (they are drawn from the same population).
- **Statistic:** A statistic is any function (*mean, variance*) of the observations in a random sample. A statistic is a random variable, and it has a probability distribution.
- **The sampling distribution of a statistic** is a probability distribution for all possible values of the ***statistic*** (*mean, variance* …) computed from a sample of size *n*. For example, the probability distribution of $\bar{X}$ is called the *sampling distribution of the mean*.
- **The sampling distribution of the sample mean $\bar{X}$** is the probability distribution of all possible values of the random variable $\bar{X}$ computed from a sample of size *n* from a population with mean μ and standard deviation σ.
- **A point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\theta}$. The statistic $\hat{\theta}$ is called the point estimator. The data from the sample is used to compute a value of a sample statistic that serves as an estimate of a population parameter.
- **Statistical inference** is concerned with making decisions about a population based on the information contained in a random sample from that population.

2.1 The mean and standard deviation of the sampling distribution of $\bar{X}$

- Suppose that a random sample of size *n* is drawn from a population with mean μ and standard deviation σ. Then, the sampling distribution of $\bar{X}$ will have mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

2.2 The Central Limit Theorem

- If $x_1, x_2, \dots, x_n$ is a random sample of size *n* taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\bar{X}$ is the sample mean, the limiting form of the distribution of

$$z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \tag{1}$$

as $n \to \infty$ $(n \geq 30)$, is the standard normal distribution *N(0, 1).*

Example 1

The results of an exam are approximately normally distributed with mean μ=100 and standard deviation σ =15. Find the probability that a random sample of size *n*=10 has mean greater than 110.

Note: $P(z \leq 2.11)=0.9826$

Solution:

We need to find P($\bar{X}$ >110).

We have $\mu_{\bar{X}} = 100$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{10}} = 4.743$

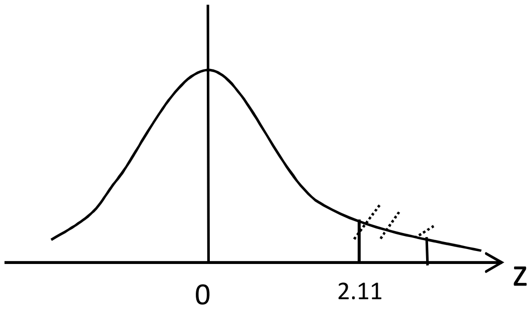
We convert $\bar{x}$ to z-score:

$Z = \frac{\bar{X}-\mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{110-100}{4.743} = 2.11$

P($\bar{X}$ >110)= P(Z > 2.11) = 1-P(Z ≤ 2.11)

From table: P(Z ≤ 2.11) = 0.9826

P($\bar{X}$ >110) = 1-0.9826 = 0.0174



Example 2

A random variable X has:

$$f_X(x) = \begin{cases} \frac{1}{2} & 4 \leq x \leq 6 \\ 0 & Ot\ erwise \end{cases}$$

Find the distribution of the sample mean of a random sample of size *n*=40.

Solution:

X has mean and variance:

$\mu_X = E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$

$= \int_4^6 \frac{1}{2} x\, dx = \frac{1}{2} . \frac{X^2}{2} |_4^6 = \frac{1}{4}(36-16) = 5$

$$\sigma_X{}^2 = E[X^2] - E[X]^2$$

$$= \int_4^6 \frac{1}{2} x^2 dx - 25 = \frac{1}{2} . \frac{1}{3} X^3 |_4^6 - 25 = 0.333$$

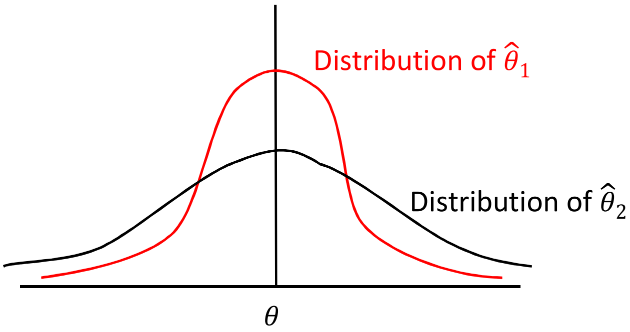
According to the *central limit theorem*, the sample distribution is approximately normal with mean $\mu_{\bar{X}} = 5$ and variance $\sigma_{\bar{X}}{}^2 = \frac{\sigma_x{}^2}{n} = \frac{0.333}{40} = 0.008\ or\ \sigma_{\bar{X}} = 0.091$

3. General concepts of point estimation

- An estimator should be close in some sense to the true value of the unknown parameter.
- The point estimator $\hat{\theta}$ is an ***unbiased*** estimator of the parameter θ if $E[\hat{\theta}] = \theta$.
- If the estimator is biased, then the difference $E[\hat{\theta}] - \theta$ is called the bias of the estimator $\hat{\theta}$.

3.1 Variance of a point estimator

- Suppose that $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimator of θ. If $\hat{\theta}_1$ has smaller variance than $\hat{\theta}_2$ then $\hat{\theta}_1$ is more likely to produce an estimate close to θ.
- The unbiased estimator with smallest variance is called the minimum variance unbiased estimator (MVUE).



- If $X_1, X_2, \ldots X_n$ is a random sample of size *n* from a normal distribution with mean μ and variance σ^2, then the sample mean $\bar{X}$ is the MVUE for μ.
- The mean squared error of an estimator $\hat{\theta}$ of the parameter θ is defined as:

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

4. Maximum likelihood estimator (MLE)

- *MLE* obtains a point estimator of a parameter by maximizing the *likelihood function*.
- If *X* is a random variable with probability distribution $f(x;\theta)$, where θ is an unknown parameter.

- Let $x_1, x_2, \dots \dots . x_n$ be the observed values in a random sample of size *n*. Then, the likelihood function $L(\theta)$ of the sample is:

$$L(\theta) = f_X(x_1;\theta) \, . \, f_X(x_2;\theta) \dots \dots . \, f_X(x_n;\theta)$$

- Note that $L(\theta)$ is now a function of only the unknown parameter θ.
- The Maximum likelihood estimator (MLE) of θ is the value of θ that maximizes $L(\theta)$.

Example 4:

Let X be a Bernoulli random variable with probability mass function:

$$f(x;P) = \begin{cases} P^x(1-P)^{1-x} & ; \ x = 0, 1 \\ 0 \quad , & otherwise \end{cases}$$

Where *P* is a parameter to be estimated.

Find the MLE $\hat{P}$ of a random sample of size *n*.

Solution:

The likelihood function is:

$$L(P) = P^{x_1}(1-P)^{1-x_1} . \ P^{x_2}(1-P)^{1-x_2} \dots \dots . . P^{x_n}(1-P)^{1-x_n}$$

$$or \ L(P) = P^{\sum_{i=1}^{n} x_i} \ (1-P)^{n-\sum_{i=1}^{n} x_i}$$

The MLE $\hat{P}$ is P that maximizes L(P) , or equivantely, lnL(P)

$$lnL(P) = \sum_{i=1}^{n} x_i \ lnP + (n - \sum_{i=1}^{n} x_i) \ln(1-P)$$

$$\frac{d \ lnL(P)}{dP} = \frac{\sum_{i=1}^{n} x_i}{\hat{P}} - \frac{n-\sum_{i=1}^{n} x_i}{1-\hat{P}} = 0 \text{ (to find the max or min)}$$

$$\Rightarrow \sum_{i=1}^{n} x_i - \hat{P} \sum_{i=1}^{n} x_i = n\hat{P} - \hat{P} \sum_{i=1}^{n} x_i$$

$$\hat{P} = \frac{\sum_{i=1}^{n} x_i}{n} \qquad \hat{P} \ is \ the \ mean \ of \ f(x;P).$$

Example 5:

Let X be a normally distributed random variable with unknown mean μ and known σ^2.

Find the MLE of μ for a random sample of size n.

Solution:

We Have: $f(x;\mu) = \frac{1}{\sqrt{2\pi}\sigma} e^{\frac{-(x-\mu)^2}{\sigma^2}}$

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

$lnL(\mu) = \ln\left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n - \sum_{i=1}^{n}\frac{(x_i-\mu)^2}{\sigma^2}$

$\frac{dlnL(\mu)}{d\mu} = -\frac{2}{\sigma^2}\sum_{i=1}^{n}(x_i - \hat{\mu}) = 0$

$\Rightarrow \sum_{i=1}^{n}(x_i - \hat{\mu}) = 0$

$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \hat{\mu} = 0$

$\sum_{i=1}^{n} x_i - \mathrm{n}\hat{\mu} = 0 \Rightarrow \hat{\mu} = \frac{\sum_{i=1}^{n} X_i}{n}$

which is the mean of the sample.