Chapter 6. Estimation, Confidence Intervals, Hypothesis Testing

6.1. Point Estimation

1. Let us begin with an example.

2. *Example.* Recall that, if $X_1 \cdots, X_n$ are a random sample of size n from the normal distribution $N(\mu, \sigma^2)$, then the sample mean $\bar{X}$ and the sample variance S^2 are independent, and

$$E(\bar{X}) = \mu, \qquad\qquad E(S^2) = \sigma^2.$$

Here, as usual,

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \text{and} \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2.$$

* In this example, μ is the population mean, σ^2 is the population variance; both μ and σ^2 are population parameters. In general, a constant associated with a population is called a population parameter. Population parameters are fixed numbers.

  $\bar{X}$ is the sample mean, S^2 is the sample variance; both $\bar{X}$ and S^2 are sample statistics. In general, a function of the sample is called a sample statistic. A sample statistic is a random variable.

  In an experiment, the observed values of $X_1, \cdots, X_n$ are denoted by $x_1, x_2, \cdots, x_n$, the observed value of $\bar{X}$ is denoted by $\bar{x}$, and the observed value of S^2 is denoted by s^2. Observed values vary from experiment to experiment.

* (example continued) Now suppose that the population parameters μ and σ^2 are unknown (fixed, but unknown), then, Since $E(\bar{X}) = \mu$, the sample mean $\bar{X}$ can be used to estimate population mean μ, and we write

$$\hat{\mu} = \bar{X}; \quad \text{(the hat on } \mu \text{ means "an estimate of")}$$

and, since $E(S^2) = \sigma^2$, the sample variance S^2 can be used to estimate population variance, and we write

$$\hat{\sigma}^2 = S^2.$$

3. Now we look at the general definition:

4. *Definition.* Suppose that the population X is a random variable that has pdf $f(x;\theta)$,

   Here, the function f is known, θ is an unknown parameter. The parameter θ is not a random variable, it has a fixed but unknown value.

   The set of all possible values of θ is called the parameter space and denoted by Ω.

   For each fixed value of θ, $f(x;\theta)$ is a pdf as a function of x. The expression $f(x;\theta)$ can be considered as a family of distributions, indexed by θ.

   We know that the distribution of X is in a certain family (the family f), we just don't known which member it is. In that sense, the distribution of X is only partially known.

To find more information about the population X, we take a sample. Let $X_1, \cdots, X_n$ be a random sample from the population X, and suppose that $x_1, \cdots, x_n$ are the observed values of the sample. We wish to find an estimation for the parameter θ.

If a function $u(X_1, \cdots, X_n)$ is used to estimate θ, then it is called an estimator of θ, and $u(x_1, \cdots, x_n)$ is called an estimate of θ.

Here, the function $u(X_1, \cdots, X_n)$ is a function of the sample, so it is a sample statistic. Since the function $u(X_1, \cdots, X_n)$ is is a function of the random variables $X_1, \cdots, X_n$, it is also a random variable. The number $u(x_1, \cdots, x_n)$ is the observed value of this sample statistic $u(X_1, \cdots, X_n)$.

Since $u(X_1, \cdots, X_n)$ is single-valued (not an interval), we say

$$u(X_1, \cdots, X_n)$$

is a point estimator.

5. *Example.* Suppose that there is a large batch of resistors. A proportion p of the resistors are defective. Here, p is unknown. We wish to make an estimate for this p. It is time-consuming to test all resistors in the batch. So we take a sample, and we choose a carefully chosen sample statistic to approximate or estimate the population parameter p.

6. There are different ways to choose a point estimator. The following definition introduces a method called the maximum likelihood estimator.

7. *Definition.* Suppose that X is a random variable, and suppose that the distribution of X belongs to the family $f(x;\theta)$ indexed by θ.

   We wish to find the value of θ that corresponds to the distribution of X. So we take a sample.

   Let $X_1, \cdots, X_n$ be a random sample from the population X. Then, $X_1, \cdots, X_n$ has joint pdf

$$F(x_1, \cdots, x_n) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta).$$

   Let $x_1, \cdots, x_n$ be the observed values of the sample.

   For these observed values $x_1, \cdots, x_n$ of the sample, define

$$L(\theta) = f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta).$$

   This $L(\theta)$ is called the likelihood function.

Note that the joint pdf $F(x_1, \cdots, x_n)$ and the likelihood function $L(\theta)$ have the same defining expression. If the expression is viewed as a function of $x_1, , \cdots, x_n$, it is the joint pdf $F(x_1, \cdots, x_n)$; on the other hand, if the expression is viewed as a function of θ, it is the likelihood function $L(\theta)$.

If $u(x_1, \cdots, x_n)$ maximizes $L(\theta)$, that is, if $L(\theta)$ is maximized at $\theta = u(x_1, \cdots, x_n)$, then $\hat{\theta} = u(X_1, \cdots, X_n)$ is called a maximum likelihood estimator (MLE) for θ, and $u(x_1, \cdots, x_n)$ is a maximum likelihood estimate.

8. We now look at an example of a MLE.

9. *Example.* Suppose that $X \sim exponential(\theta)$. Suppose that θ is unknown. Then, X has pdf

$$f(x;\theta) = \theta^{-1} e^{-x/\theta}, \quad x > 0.$$

The parameter space is $\Omega = (0, \infty)$.

Let $X_1, \cdots, X_n$ be a random sample from the population X. And, $X_1, \cdots, X_n$ has joint pdf

$$\begin{aligned} F(x_1, \cdots, x_n) &= f(x_1;\theta) f(x_2;\theta) \cdots f(x_n;\theta) \\ &= \theta^{-1} e^{-x_1/\theta} \cdot \cdots \cdot \theta^{-1} e^{-x_n/\theta} \\ &= \theta^{-n} e^{-(x_1+x_2+\cdots+x_n)/\theta}. \end{aligned}$$

Let $t = \sum_{i=1}^{n} x_i$, then

$$F(x_1, \cdots, x_n) = \theta^{-n} e^{-t/\theta}.$$

Hence, the likelihood function is

$$L(\theta) = \theta^{-n} e^{-t/\theta}.$$

After some derivation based on Calculus (see remark below), it can be shown that L achieves its maximum at $\theta = t/n$. Hence,

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

is the maximum likelihood estimator (MLE) of θ. Since

$$E(\hat{\theta}) = \theta,$$

$\hat{\theta}$ is an unbiased estimator for θ.

10. *Remark.* We now show that the function

$$L(\theta) = \theta^{-n} e^{-t/\theta}, \quad t > 0,$$

achieves its maximum at $\theta = t/n$. Calculation shows that

$$L'(\theta) = \frac{n e^{-t/\theta}}{\theta^{n+2}} (t/n - \theta).$$

So, if $\theta > t/n$, then $L'(\theta) < 0$ and $L(\theta)$ is decreasing; if $0 < \theta < t/n$, then $L'(\theta) > 0$ and $L(\theta)$ is increasing. So, $L(\theta)$ achieves its maximum at $\theta = t/n$.