Chapter 3: Point Estimation

In this chapter, we being by formally outlining the purpose of statistical inference. We follow this by discussing the problem of point estimation of population parameters. We confine our formal developments of specific estimation procedures to problems involving one sample.

Statistical Inference:

Statistical inference consists of those methods by which one makes inferences or generalizations about a population. There are two types of methods, the **classic method** of estimating a population parameter, whereby inferences are based strictly on information obtained from a random sample selected from the population, and the **Bayesian method**, which utilizes prior subjective knowledge about the probability distribution of the unknown parameters in conjunction with the information provided by the sample data. Throughout of this chapter and the next, we shall use classical methods to estimate unknown population parameters such as the mean and the variance by computing statistics from random samples and applying the theory of sampling distributions, much of which was covered in Chapter 2. Bayesian estimation will be discussed in Chapter 5.

Statistical inference may be divided into two major areas: **estimation** and **tests of hypotheses**, see Figure 3.1. We treat only estimation area in this course. Estimation methods divide into two parts, **point estimation** which we will discuss it in this chapter and **interval estimation** that will discuss in Chapter 4.

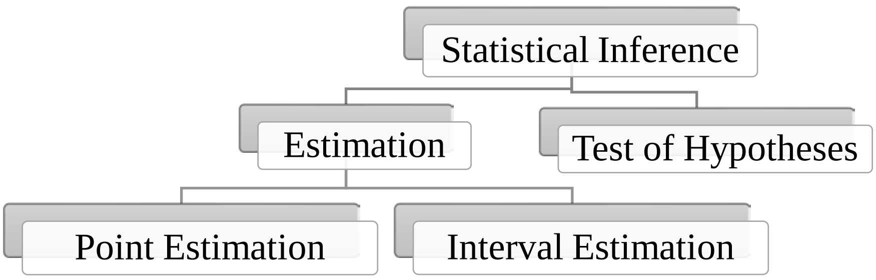

Figure 3.1

Point Estimate and Estimator:

A **point estimate** of some population parameter θ is a single value $\hat{\theta}$ of an **estimator** which is a statistic T. For example, the value $\bar{x}$ of the estimator

(statistic) $\bar{X}$, computed from a sample of size n is a point estimate of the population parameter μ.

3.1 Point Estimation Methods

This section introduced two different methods to derive the point estimator that are, method of moments estimator (MME) and maximum likelihood estimator (MLE).

3.1.1 Method of Moments Estimation

Let $X_1, X_2, \ldots, X_n$ be random sample of size n from a distribution with probability distribution $f(x;\ \theta_1, \theta_2, \ldots, \theta_r), (\theta_1, \ldots, \theta_r) \in \Omega$. The expectation $\mu_k' = E(X^k)$ is frequently called the ***k*th moment of the distribution**, $k = 1, 2, 3, \ldots$. The sum $M_k = \sum_{i=1}^{n} \frac{X_i^k}{n}$ is the ***k*th moment of the sample**, $k = 1, 2, 3, \ldots$. The **method of moments estimators,** $\tilde{\theta}_1, \tilde{\theta}_2, \ldots, \tilde{\theta}_r$, are then the solution of the following *r*th equations,

$$\mu_i' = M_i$$

for $\theta_1, \theta_2, \ldots, \theta_r$, $i = 1, 2, \ldots, r$.

3.1.2 Maximum Likelihood Estimation

Maximum likelihood estimation is one of the most important approaches to estimation in all of statistical inference. In this section we develop statistical inference (point estimation) based on likelihood methods. We show that this procedure are asymptotically optimal under certain conditions (regularity conditions).

Likelihood Function

Suppose that $X_1, \ldots, X_n$ are independent identically distributed (iid) random variables with common probability density function (continuous case) or probability mass function (discrete case), $f(x; \theta)$. Then, the likelihood function is given by,

$$L(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta), \theta \in \Omega.$$

where $x = (x_1, \dots \dots, x_n)$. Because we will treat L as a function of θ in this section, we will often write it as $L(\theta)$. Actually, the log or ln of this function is usually more convenient to work with mathematically. Denote the $\log L(\theta)$ by

$$\log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta), \theta \in \Omega.$$

Note that there is no loss of information in using $\log L(\theta)$ because the log is a one-to-one function. In this section, we will generally consider X as a random variable.

Maximum Likelihood Estimator:

Given independent observations $x_1, x_2, \dots .., x_n$ from a probability distribution $f(x;\ \theta_1, \theta_2, \dots .., \theta_r), (\theta_1, \dots, \theta_r) \in \Omega$, the maximum likelihood estimators $\hat{\theta}_1, \hat{\theta}_2, \dots .., \hat{\theta}_r$ are that which maximizes the likelihood function $L(\theta_1, \theta_2, \dots .., \theta_r; x)$.

To determine the MLE, we use the following **estimating equations** (EE). Then, the MLE is the solution of these equations

$$\frac{\partial L(\theta_i; x)}{\partial \theta_i} = 0 \ \text{ or } \ \frac{\partial \log\ L(\theta_i; x)}{\partial \theta_i} = 0, \quad i = 1,2, .., r$$

There is no guarantee that the MLE exists or if it does whether it is unique.

Example 3.1:

Consider a Poisson distribution with probability mas function

$$f(x, \mu) = \frac{e^{-\mu} \mu^x}{x!}, x = 0, 1, 2, \dots.$$

Supposed that a random sample $X_1, X_2, \dots, X_n$ is taken from the distribution. Find:

1. The method of moments estimator of μ.
2. The maximum likelihood estimator of μ.

Solution:

1. Since the Poisson distribution has one parameter, then we will derive only the first moment of the distribution and the first moment of the sample, as following

$$E(X) = \mu \text{ and } M_1 = \sum_{i=1}^{n} \frac{X_i}{n}$$

Solving the equation, $E(X) = M_1$, then the MME is obtained as

$$\tilde{\mu} = \sum_{i=1}^{n} \frac{X_i}{n} = \bar{X}.$$

2. The likelihood function is

$$L(x_1, x_2, \dots . x_n; \mu) = \prod_{i=1}^{n} f\,(x_i, \mu) = \frac{e^{-n\mu} \mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

Now consider

$$\log L(x_1, x_2, \dots . x_n; \mu) = -n\mu + \sum_{i=1}^{n} x_i \log\, \mu - \log\, \prod_{i=1}^{n} x_i\,!,$$

$$\frac{\partial \log\, L(x_1, x_2, \dots, x_n; \mu)}{\partial_\mu} = -n + \sum_{i=1}^{n} \frac{x_i}{\mu} = 0,$$

Solving for $\hat{\mu}$, the maximum likelihood estimator is given by

$$\hat{\mu} = \sum_{i=1}^{n} \frac{X_i}{n} = \bar{X}.$$

The second derivative of the log-likelihood function is negative, which implies that the solution above indeed is maximum. Since μ is the mean of the Poisson distribution (Chapter 1), the sample average would certainly seem like a reasonable estimator.

Example 3.2:

Suppose 10 rats are used in a biomedical study where they are injected with cancer cells and then given a cancer drug that is designed to increase their survival rate. The survival times, in months, are 14, 17, 27, 18, 12, 8, 22, 13, 19, and 12. Assume that the exponential distribution applies.

$$f(x, \beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}.$$

Drive the method of moments and the maximum likelihood estimates of the mean survival time.

Solution:

To find the method of moments estimate we need to calculate the following moments

$$E(X) = \beta \text{ and } M_1 = \sum_{i=1}^{10} \frac{X_i}{n}$$

By equating these moments, we get the MME as

$$\tilde{\beta} = \sum_{i=1}^{10} \frac{X_i}{n} = \bar{X} = 16.2.$$

Now, the log-likelihood function for the date, given $n = 10$, is

$$\log L(x_1, x_2, \ldots, x_{10}; \beta) = -10 \log \beta - \frac{1}{\beta} \sum_{i=1}^{10} X_i,$$

Setting

$$\frac{\partial \log L}{\partial \beta} = -\frac{10}{\beta} + \frac{1}{\beta^2} \sum_{i=1}^{10} X_i = 0,$$

Applies that

$$\hat{\beta} = \bar{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = 16.2.$$

As a result, the estimator of the parameter β, the population mean, is the sample average $\bar{X}$.

3.2 Properties of the Estimators

In this section, we will study several measures of the quality of an estimator, so that we can choose the best. Some of these measures tell us the quality of the estimator with small samples, while other measures tell us the quality of the estimator with large samples. The last are also known as asymptotic properties of estimators.

Small-sample Properties: (*n* finite or infinite)	**Large-sample Properties: (*n*→∞)**
Unbiasedness (mean).	Asymptotic unbiasedness
Sufficiency	Consistency.
Complete	Asymptotic efficiency
Efficiency (variance).	Asymptotic normality.

3.2.1 Unbiasedness

Let $X_1, \ldots, X_n$ be a random sample from the probability distribution $f(x;\theta)$; and let T denote an estimator of θ. We say that a statistic T is an **unbiased** estimator of θ if

$$E(T) = \theta, \qquad \forall\, \theta$$

If T is not unbiased (that is, $E(T) \neq \theta$), we say that T is a **biased** estimator of θ.

3.2.2 Mean Squared Error

Let $X_1, \ldots, X_n$ be a random sample from the probability distribution $f(x;\theta)$. Let a statistic T is an estimator of θ. Then, the mean squared error of T, MSE, is given by

$$MSE(T) = E[(T-\theta)^2] = Var(T) - \left(\theta - E(T)\right)^2$$

The term $\left(\theta - E(T)\right)$ is called the bias of the estimator T. Note That if T is an unbiased estimator of θ, then the MSE is

$$MSE(T) = Var(T)$$

Proof:

$$MSE(T) = E[(T-\theta)^2] = E\left[\left(\left(T - E(T)\right) - \left(\theta - E(T)\right)\right)^2\right]$$

$$= E\left[\left(T-E(T)\right)^2 - 2\left(T-E(T)\right)\left(\theta - E(T)\right) + \left(\theta - E(T)\right)^2\right]$$

$$= E\left(T-E(T)\right)^2 - 2E\left(T-E(T)\right)\left(\theta - E(T)\right) + E\left(\theta - E(T)\right)^2$$

$$= Var(T) - \left[\left(\theta - E(T)\right)^2\right]$$

Theorem 3.1:

If T_1 and T_2 are two estimators of θ, then T_1 is better estimator than T_2 if

$$MSE(T_1) \leq MSE(T_2).$$

3.2.3 Consistency

Any estimator (statistic) T that converges to a parameter θ is called a **consistent** estimator of that parameter θ, i.e.

$$\lim_{n\to\infty} P(|T-\theta| \geq \varepsilon) = 0, \ \ \forall\theta.$$

Theorem 3.2:

An estimator T_n based on a sample of size *n* is consistent for θ if

1. $\lim_{n\to\infty} E(T_n) = \theta$ (asymptotically unbiased)

and 2. $\lim_{n\to\infty} Var(T_n) = 0.$

3.2.4 Sufficiency

Let $X_1, X_2, \ldots., X_n$ denote a random sample of size n from a distribution $f(x;\ \theta), \theta \in \Omega$. Let $T(x)$ be a statistic whose distribution is $f_T(t;\ \theta)$. Then, T is a **sufficient statistic** of θ if and only if

$$\frac{\prod_{i=1}^{n} f(x_i;\theta)}{f_T(t;\theta)} \text{ does not depend on } \theta.$$

Theorem 3.3: (Factorization Theorem)

Let $X_1, X_2, \ldots., X_n$ denote a random sample from a distribution $f(x;\ \theta), \theta \in \Omega$. The statistic $T(x)$ is a sufficient statistic of θ if and only if we can find two nonnegative functions, K_1 and K_2, such that

$$\prod_{i=1}^{n} f(x_i;\ \theta) = K_1(t,\theta).K_2(x_1, x_2, \ldots., x_n),$$

where $K_2(x_1, x_2, \ldots., x_n)$ does not depend upon θ.

Theorem 3.4:

Let $X_1, X_2, \ldots., X_n$ denote a random sample from a distribution that has probability distribution $f(x;\ \theta).\theta \in \Omega$. If a sufficient statistic $T(x)$ of θ exist and if a maximum likelihood estimator $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of $T(x)$.

Example 3.3:

Let $X_1, X_2, \dots., X_n$ be a random sample has exponential distribution with parameter β as following:

$$f(x,\beta) = \frac{1}{\beta} e^{-x/\beta}, \ \ x > 0$$

Show that the estimator $\bar{X}$ is an unbiased, consistent and sufficient statistic estimator, then find the mean squared error of β.

Solution:

Remember: $\int_0^\infty x^{\alpha-1} e^{-\frac{x}{\beta}} dx = \Gamma(\alpha)\,\beta^\alpha$.

First, we have to find the mean and the variance of X

$$E(X) = \int_0^\infty \frac{1}{\beta} x\, e^{-x/\beta} dx = \frac{1}{\beta}\Gamma(2)\beta^2 = \beta$$

$$E(X^2) = \int_0^\infty \frac{1}{\beta} x^2 e^{-x/\beta} dx = \frac{1}{\beta}\Gamma(3)\beta^3 = 2\beta^2$$

$$Var(X) = E(X^2) - \big(E(X)\big)^2 = 2\beta^2 - \beta^2 = \beta^2$$

Then,

$$E(\bar{X}\,) = E\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^n E(X_i) = \frac{1}{n}(n\beta) = \beta.$$

Thus, the statistic $\bar{X}$ is an unbiased estimator of β. Now, we will find the variance of $\bar{X}$ as

$$Var(\bar{X}\,) = Var\left(\frac{\sum_{i=1}^n X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^n Var(X_i) = \frac{1}{n^2}(n\beta^2) = \frac{\beta^2}{n}.$$

Thus,

$$\lim_{n\to\infty} Var(\bar{X}\,) = \lim_{n\to\infty}\left(\frac{\beta^2}{n}\right) = 0,$$

Therefore, since $\bar{X}$ is an unbiased estimator of β and $\lim_{n\to\infty} Var(\bar{X}\,) = 0$, from Theorem 3.2, the estimator $\bar{X}$ is a consistent estimator.

Now, we need to derived the distribution of T which can be found by using the mgf transformation method as

$$M_{\bar{X}}(t) = E\left(e^{\bar{X}t}\right) = E\left(e^{\frac{\sum_{i=1}^{n} x_i}{n}t}\right) = \left(M_{X_i}\left(\frac{t}{n}\right)\right)^n = \left(1 - \frac{\beta}{n}t\right)^{-n}$$

which is the mgf of $Gamma\left(n, \frac{\beta}{n}\right)$, thus the pdf of $\bar{X}$ is (let $T = \bar{X}$)

$$f_T\left(t;\ n, \frac{\beta}{n}\right) = \frac{n^n}{\beta^n \Gamma(n)} t^{n-1} e^{-\frac{nt}{\beta}}\ ,\ \ t > 0,$$

$$\prod_{i=1}^{n} f(x_i;\ \beta) = \prod_{i=1}^{n} \frac{1}{\beta} e^{-x_i/\beta} = \frac{1}{\beta^n} e^{-\sum_{i=1}^{n} x_i/\beta} \quad ,$$

$$\frac{\prod_{i=1}^{n} f(x_i; \beta)}{f_T\left(t; n, \frac{\beta}{n}\right)} = \frac{\frac{1}{\beta^n} e^{-\sum_{i=1}^{n} x_i/\beta}}{\frac{n^n}{\beta^n \Gamma(n)} t^{n-1} e^{-\frac{nt}{\beta}}} = \frac{\Gamma(n)}{n^n} t^{1-n}$$

which does not depend on β, thus we conclude that $T = \bar{X}$ is a sufficient statistic estimator.

The MSE of $\bar{X}$ is given by

$$MSE(\bar{X}\ ) = Var(\bar{X}\ ) = \frac{\beta^2}{n} \text{ (since } \bar{X} \text{ is an unbiased estimator).}$$

Thus, the estimator $\bar{X}$ is unbiased, consistent and sufficient statistic estimator of β. Notice that the estimator $\bar{X}$ is the MME and the MLE of β.

Example 3.4:

Let $X_1, X_2, \dots, X_n$ be a random sample with Poisson pmf and parameter μ, i.e.

$$f(x, \mu) = \frac{e^{-\mu}\mu^x}{x!}, x = 0, 1, 2, \dots.$$

Show that the MLE of μ is an unbiased, consistent and sufficient statistic estimator then find the mean squared error of μ.

Solution:

From example 3.1, the MLE of μ is $\bar{X}$. First, we want to find the mean and the variance of $\bar{X}$:

We know that the mean and the variance of Poisson distribution with parameter μ are given by

$$E(X) = Var(X) = \mu$$

Then,

$$E(\bar{X}) = E\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}(n\mu) = \mu.$$

which conclude that the MLE of μ is an unbiased estimator. Thus,

$$MSE(\bar{X}) = Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{1}{n^2}(n\mu) = \frac{\mu}{n}$$

$$\lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty}\left(\frac{\mu}{n}\right) = 0,$$

Therefore, the estimator $\bar{X}$ is a consistent estimator of μ.

Now,

$$\prod_{i=1}^{n} f(x_i,\mu) = \prod_{i=1}^{n} \frac{e^{-\mu}\mu^{x_i}}{x_i!} = \frac{e^{-n\mu}\mu^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!} = \frac{e^{-n\mu}\mu^{n\bar{x}}}{\prod_{i=1}^{n} x_i!}$$

Thus, $\prod_{i=1}^{n} f(x_i,\mu)$ can be written by a product of two functions $K_1(t,\theta) = e^{-n\mu}\mu^{n\bar{x}}$ which depends on the parameter μ and the MLE, $T = \bar{x}$ and $K_2(x_1, x_2, \dots., x_n) = \frac{1}{\prod_{i=1}^{n} x_i!}$ which depends only on the random sample. Therefore, we conclude that $T = \bar{X}$ is a sufficient statistic estimator.

Thus, the MLE, $\bar{X}$ is unbiased, consistent and sufficient statistic estimator of μ.

Theorem 3.5:

Let $X_1, X_2, \dots., X_n$ denote a random sample from a distribution $f\left(x;\, \acute{\theta}\right), \acute{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Then, statistic $\acute{T} = (T_1, T_2, \dots., T_k)$ are **joint sufficient statistic** of $\acute{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$ if and only if

$$L(x;\acute{\theta}) = \prod_{i=1}^{n} f(x_i;\, \acute{\theta}) = K_1(\acute{t},\acute{\theta}).K_2(x_1, x_2, \dots., x_n),$$

where $K_2(x_1, x_2, \dots., x_n)$ does not depend on θ.

Example 3.5:

Let $X_1, X_2, \dots, X_n$ be a random sample drawn from continuous uniform distribution when $x \in (0,\theta)$. Find the following:

(a) The MLE of θ.
(b) Prove that $Y_n = \text{Maximum}(X_1, X_2, \dots, X_n)$ is a sufficient statistic, asymptotically unbiased and consistent estimator of θ.
(c) An unbiased estimator of θ.

Solution:

(a) The pmf and cdf of the uniform distribution of $x \in (0, \theta)$ are defined as

$$f(x,\theta) = \frac{1}{\theta} \quad \text{and} \quad F(x) = \frac{x}{\theta}$$

and the likelihood function is given by

$$L(x_1, x_2, \dots, x_n;\ \theta) = \frac{1}{\theta^n}\ ,\ \ 0 < x_i \le \theta$$

Then, the maximum of such functions cannot be found by differentiation but by selecting θ as small as possible. Now, each $x_i \le \theta$, in particular $Y_n \le \theta$. Thus, the likelihood function attains to the maximum value when

$$L\left(x_1, x_2, \dots, x_n; \hat{\theta}\right) = \frac{1}{(Y_n)^n}$$

or $\hat{\theta} = Y_n$ is the MLE for θ.

(b) To find the properties of the estimator Y_n, we should first derive the distribution of it as:

$$f(y_n, \theta) = \frac{n}{\theta^n} y_n^{n-1}\ \ ,\ \ \ 0 < y_n \le \theta$$

Thus,

$$\frac{\prod_{i=1}^{n} f(x_i;\theta)}{f\left(y_n,\theta\right)} = \frac{1/\theta^n}{(n/\theta^n)\ y^{n-1}} = \frac{1}{n\ y_n^{n-1}} \text{ dose not depend on } \theta$$

The estimator Y_n is sufficient statistic for θ. Now, the mean and the variance of are given by

$$E\left(y_n\right) = \int_0^\theta \frac{n}{\theta^n} y_n^n\, dy = \frac{n}{\theta^n(n+1)} y_n^{n+1}\Big|_0^\theta = \frac{n\,\theta}{(n+1)}$$

$$E\left(y_n^2\right) = \int_0^\theta \frac{n}{\theta^n} y_n^{n+1}\, dy = \frac{n}{\theta^n(n+2)} y_n^{n+2}\Big|_0^\theta = \frac{n\,\theta^2}{(n+2)}$$

$$Var(y_n) = E(y_n^2) - \left(E(y_n)\right)^2$$

$$= \frac{n\,\theta^2}{(n+2)} - \frac{n^2\,\theta^2}{(n+1)^2} = \frac{n\,\theta^2}{(n+2)(n+1)^2}$$

Thus,

$$\lim_{n\to\infty} E(y_n\,) \frac{n\,\theta}{(n+1)} = \lim_{n\to\infty} \frac{n\,\theta}{(n+1)} = \theta$$

$$\lim_{n\to\infty} Var(y_n\,) = \lim_{n\to\infty} \frac{n\,\theta^2}{(n+2)(n+1)^2} = 0$$

Therefore, Y_n is asymptotically unbiased and consistent estimator of θ.

(c) Since $E(y_n) = \frac{n\,\theta}{(n+1)}$, thus we can choose $T = \frac{(n+1)}{n} y_n$ which is an unbiased estimator for θ such that $E(T) = E\left(\frac{(n+1)}{n} y_n\right) = \theta.$

Example 3.6:

Let $X_1, X_2, \ldots\ldots, X_n$ denote a random sample from a distribution that is $N(\mu, \sigma^2), -\infty < \mu < \infty,\ \sigma^2 > 0$. Find the following:

1. Maximum likelihood estimators of μ and σ^2.
2. Method of moments estimators of μ and σ^2.
3. Properties of MLE and MME of μ and σ^2.

Solution:

1. Maximum likelihood estimators of μ and σ^2:

The pdf of the $N(\mu, \sigma^2)$ is

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x > 0$$

The likelihood and the logarithm of the likelihood function may be written in the form

$$L(\mu, \sigma^2; x_1, \ldots\ldots, x_n) = \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

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

$$\log L(\mu, \sigma^2 ; x_1, \dots\dots, x_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2,$$

We observe that we may maximum by differentiation $\ln L(\mu, \sigma^2 ; x_1, \dots\dots, x_n)$ with respect to μ and σ^2. We have

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

$$\frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2,$$

If we equate these partial derivatives to zero and solve simultaneously the two equations thus obtained, the solutions for μ and σ^2 are found to be

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0 \Rightarrow \sum_{i=1}^{n} x_i - n\mu = 0 \Rightarrow \hat{\mu} = \bar{X},$$

$$-\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0 \Rightarrow \sum_{i=1}^{n}(x_i - \mu)^2 = n\sigma^2,$$

$$\Rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \hat{\mu})^2}{n} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n},$$

2. The method of moments estimators of μ and σ^2:

Since we want to find MME for two parameters μ and σ^2, then we have to equate first two population moments

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

with first two sample moments

$$M_1 = \sum_{i=1}^{n}\frac{X_i}{n}, M_2 = \sum_{i=1}^{n}\frac{X_i^2}{n},$$

Then, we get

$$\tilde{\mu} = \bar{X} \text{ , and}$$

$$\sigma^2 + \mu^2 = \sum_{i=1}^{n}\frac{X_i^2}{n} \Rightarrow \tilde{\sigma}^2 = \sum_{i=1}^{n}\frac{X_i^2}{n} - \left(\sum_{i=1}^{n}\frac{X_i}{n}\right)^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n},$$

Which are equal to the MLEs of μ and σ^2.

3. Estimators properties:

a) Unbiasedness:

$$E(\hat{\mu}) = E(\bar{X}) = \mu$$

Thus, the estimator $\bar{X}$ is an unbiased estimator of μ.

$$E(\hat{\sigma}^2) = E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right),$$

We know that the term $\sum_{i=1}^{n}(X_i - \bar{X})^2$ can be written as

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2,$$

Then,

$$E(\hat{\sigma}^2) = \frac{1}{n}\left[\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right],$$

$$= \frac{1}{n}\left(\sum_{i=1}^{n} \sigma^2 - nVar(\bar{X})\right)$$

$$= \frac{1}{n}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \frac{(n-1)\sigma^2}{n}$$

Therefore, $\hat{\sigma}^2$ is biased estimator of σ^2.

Note: The estimator $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an unbiased estimator (**Prove**).

b) Mean squared error:

The MSE of μ and σ^2 are given, respectively, by

$$MSE(\hat{\mu}) = MSE(\bar{X}) = Var(\bar{X}) = \frac{\sigma^2}{n}$$

$$MSE(\hat{\sigma}^2) = Var(\hat{\sigma}^2) + E\left[\left(\sigma^2 - E(\hat{\sigma}^2)\right)^2\right]$$

We need to find the variance of $\hat{\sigma}^2$. From Theorem 2.8,

$$\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Define $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$, now since $X \sim N(\mu, \sigma^2)$, thus we can conclude that

$$\frac{n\,S_1^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Therefore,

$$Var\left(\frac{n\,S_1^2}{\sigma^2}\right) = 2(n-1) \Rightarrow Var(S_1^2) = \frac{2(n-1)\sigma^4}{n^2}$$

The MSE is, then given by

$$MSE(\hat{\sigma}^2) = \frac{2(n-1)\sigma^4}{n^2} + \left(\sigma^2 - \frac{(n-1)\sigma^2}{n}\right)^2$$

$$= \frac{2(n-1)\sigma^4}{n^2} + \left(\frac{\sigma^2}{n}\right)^2 = \frac{(2n-1)\sigma^4}{n^2}$$

b) Consistency:

The estimator $\bar{X}$ is a consistent estimator of μ because

1. It is an unbiased estimator of μ.
2. $\lim_{n\to\infty} Var(\bar{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0$.

The estimator $\hat{\sigma}^2$ of σ^2 is also consistent estimator because

1. $\lim_{n\to\infty} E(\hat{\sigma}^2) = \lim_{n\to\infty} \left[\frac{(n-1)\sigma^2}{n}\right] = \lim_{n\to\infty} \left[\sigma^2 - \frac{\sigma^2}{n}\right] = \sigma^2$ (asymptotically unbiased).
2. $\lim_{n\to\infty} Var(\hat{\sigma}^2) = \lim_{n\to\infty} \left[\frac{(2n-1)\sigma^4}{n^2}\right] = \lim_{n\to\infty} \left[\frac{2\sigma^4}{n} - \frac{\sigma^2}{n^2}\right] = 0$.

d) Sufficiency:

The likelihood function of $N(\mu, \sigma^2)$ is obtained as

$$\prod_{i=1}^{n} f(x_i;\, \mu, \sigma^2) = \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-\mu)^2}$$

$$= \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(X_i^2-2\mu X_i+\mu^2\right)}$$

$$= \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} X_i^2-2\mu\sum_{i=1}^{n} X_i+n\mu^2\right]}$$

Let $T_1 = \sum_{i=1}^{n} X_i$, $T_2 = \sum_{i=1}^{n} X_i^2$. Then, we can write

$$\prod_{i=1}^{n} f(x_i;\, \mu, \sigma^2) = K_1(T_1, T_2;\, \mu, \sigma^2).K_2(X)$$

where $K_1(T_1, T_2, \mu, \sigma^2) = \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}[T_2-2\mu T_1+n\mu^2]}$ and $K_2(X) = 1$. Therefore, (T_1, T_2) are jointly sufficient statistic of (μ, σ^2).

Exponential Family:

A probability distribution $f(x, \theta)$ is said to be a member of the exponential family if it can be written of the form

$$f(x, \theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

where, 1. $a(\theta)$ and $c(\theta)$ are functions of parameter θ.

2. $b(x)$ and $d(x)$ are functions of the random sample X.

Example 3.7:

If $X_1, X_2, \dots, X_n$ is a random sample, determine whether the following probability distribution are member of exponential family or not:

1. $Expnential\left(\frac{1}{\theta}\right)$.
2. $Bernoulli(p)$.

Solution:

1. The pdf of the exponential distribution with parameter $\frac{1}{\theta}$ is defined as

$$f(x;\theta) = \theta e^{-\theta x}, \quad x \geq 0$$

It is a member of exponential family where $a(\theta) = \theta, b(x) = 1, c(\theta) = -\theta, d(x) = x$.

2. The pmf of Bernoulli distribution with parameter p is

$$f(x;p) = p^x q^{1-x}, \; x = 0, 1.$$

which can be written as

$$f(x;p) = e^{x\, Lnp} e^{(1-x)\, Lnq} = e^{Lnq\, +\, x\, (Lnp - Lnq)}$$

Therefore, the Bernoulli distribution is a member of exponential family where $a(p) = e^{Lnq}, b(x) = 1, c(p) = Lnp - Lnq, d(x) = x$.

3.2.5 Minimal Sufficiency

A sufficient statistic T is a minimal sufficient statistic if, for any other sufficient statistic U, T is a function of U.

Theorem 3.6:

If $X_1, X_2, \dots, X_n$be random sample with probability distribution $f(x,\theta)$ and let $T(x)$ be a statistic of the random sample. Suppose for any random sample

$Y_1, Y_2, \dots, Y_n$ from probability distribution $f(y,\theta)$ such that $T(y)$ is a statistic and the ratio

$$\frac{\prod_{i=1}^{n} f(x_i,\theta)}{\prod_{i=1}^{n} f(y_i,\theta)} \text{ does not depend on } \theta \text{ if and only if } T(x) = T(y).$$

Then, $T(x)$ is a **minimal sufficient statistic** estimator of θ.

Example 3.8:

If $X_1, X_2, \dots, X_n$ are independent identically random sample from $Poisson(\theta)$. Show that $T = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for θ.

Solution:

The pmf of $Poisson(\theta)$ is given as

$$f(x,\theta) = \frac{e^{-\theta}\theta^{x}}{x!}, x = 0, 1, 2, \dots.$$

Then, for any random sample $Y \sim Poisson(\theta)$

$$\frac{\prod_{i=1}^{n} f(x_i,\theta)}{\prod_{i=1}^{n} f(y_i,\theta)} = \frac{e^{-n\,\theta}\,\theta^{\sum_{i=1}^{n} x_i} / \prod_{i=1}^{n} x_i!}{e^{-n\,\theta}\,\theta^{\sum_{i=1}^{n} y_i} / \prod_{i=1}^{n} y_i!} = \frac{\theta^{\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i}}{\prod_{i=1}^{n} x_i! / \prod_{i=1}^{n} y_i!},$$

which does not depend on θ iff $\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i$. This implies that $T = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for θ.

Theorem 3.7:

If $X_1, X_2, \dots, X_n$ be random sample from exponential family,

$$f(x,\theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

Then, $T = \sum_{i=1}^{n} d(x_i)$ is a **minimal sufficient statistic** estimator of θ.

Theorem 3.8:

If $X_1, X_2, \dots, X_n$ be a random sample from

$$f(x,\acute{\theta}) = a(\acute{\theta})b(x)e^{\sum_{j=1}^{k} c_j(\acute{\theta})d_j(x)}$$

where $\acute{\theta}$ vector of parameters, $\acute{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$. Then,

$$T_j = \sum_{i=1}^{n} d_j(x_i), \quad j = 1, 2, \dots, k;$$

are **jointly minimal sufficient statistic** estimators of $\acute{\theta} = (\theta_1, \theta_2, \dots, \theta_k)$.

Example 3.9:

Find a minimal sufficient statistic for the probability distribution in Example 3.6.

Solution:

Since $d(x) = x$ for the exponential and Bernoulli distributions, then the statistic $T = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for both distributions.

3.2.6 Completeness

A sufficient statistic $T(x)$ of θ is called complete if for any function $g(T)$ such that

$$E\big(g(T)\big) = 0, \text{ for all } \theta \text{ implies that } g(T) = 0, \text{ for all } T.$$

Theorem 3.9:

Let $X_1, X_2, \dots, X_n$ be a random sample from $f(x, \theta)$ such that

$$f(x, \theta) = a(\theta) b(x) e^{c(\theta) d(x)}$$

Then, $T = \sum_{i=1}^{n} d(x_i)$ is **complete minimal sufficient statistic** of θ.

Examples 3.10:

Let $X_1, X_2, \dots, X_n$ be a random sample from $Bernoulli(p)$. Show that $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for p.

Solution:

From Example 3.6, we found that Bernoulli distribution is a member of exponential family with $d(x) = x$. Therefore, by using Theorem 3.9, $T = \sum_{i=1}^{n} X_i$ is complete minimal sufficient statistic for p.

Now, we want to use the definition of completeness to get the same result:

Since $X \sim Bernoulli(p)$, then

$$M_T(s) = E(e^{ts}) = E\left(e^{s\sum_{i=1}^{n} X_i}\right) = \left(M_{X_1}(s)\right)^n = (q + pe^t)^n,$$

which is the mgf of $Binomial(n, p)$. Thus, the pdf of T is

$$f(t) = \binom{n}{t} p^t q^{n-t}, t = 0, 1, \dots, n$$

Suppose for any function of T , $g(T)$, that

$$E\big(g(T)\big) = \sum_{t=0}^{n} g(T) \tbinom{n}{t} p^t q^{n-t} = q^n \sum_{t=0}^{n} g(T) \tbinom{n}{t} \left(\tfrac{p}{q}\right)^t = 0$$

$$\Rightarrow g(0) \binom{n}{0} \left(\frac{p}{q}\right)^0 + g(1) \binom{n}{1} \left(\frac{p}{q}\right) + \cdots + g(n) \binom{n}{n} \left(\frac{p}{q}\right)^n = 0$$

$$\Rightarrow g(0) = g(1) = \cdots = g(n) = 0$$

$$\Rightarrow g(T) = 0, \text{ for all } T.$$

Thus, T is complete sufficient statistic for p.

Example 3.11:

Let $X_1, X_2, \dots, X_n \sim \theta e^{-\theta x}, x \geq 0$. Show that $T = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ.

Solution:

Since $X \sim Exponential\left(\frac{1}{\theta}\right)$, then the distribution of $T = \sum_{i=1}^{n} X_i$ is given as

$$M_T(s) = E(e^{ts}) = E\left(e^{s\sum_{i=1}^{n} x_i}\right) = \left(M_{X_1}(s)\right)^n = \left(\frac{\theta}{\theta - t}\right)^n,$$

which is the mgf of $Gammma\left(n, \frac{1}{\theta}\right)$. Thus, the pdf of T is

$$f_T(t) = \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t}, t > 0$$

Then, $$E\big(g(T)\big) = \int_0^\infty g(t) \frac{\theta^n}{\Gamma(n)} t^{n-1} e^{-\theta t} dt = 0$$

Only $g(t)\frac{\theta^n}{\Gamma(n)}t^{n-1}=0 \Leftrightarrow g(t)=0$, for all T.

Therefore, T is complete sufficient statistic for θ.

Score Function

Let $X_1, X_2, \ldots, X_n$ be a random sample from probability distribution $f(x,\theta)$, then the score function, $u(\theta)$, is the derivative of the log-likelihood function with respect to the parameter θ:

$$u(\theta)=\frac{\partial}{\partial\theta}\log L(x,\theta)$$

Properties of Score Function:

1. **Mean**

$$\boldsymbol{E[u(\theta)]=0}$$

Proof:

$$E[u(\theta)]=E\left[\frac{\partial}{\partial\theta}\log L(x,\theta)\right]$$

$$=\int_{x_1}\ldots\int_{x_n}L(x,\theta)\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)\,dx_1\ldots dx_n$$

$$=\int_{x_1}\ldots\int_{x_n}L(x,\theta)\left(\frac{\frac{\partial L(x,\theta)}{\partial\theta}}{L(x,\theta)}\right)\,dx_1\ldots dx_n$$

$$=\frac{\partial}{\partial\theta}\int_{x_1}\ldots\int_{x_n}L(x,\theta)\,dx_1\ldots dx_n=\frac{\partial}{\partial\theta}(1)=0$$

2. **Variance (Fisher Information)**

$$\boldsymbol{Var[u(\theta)]=E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right]}$$

Proof:

$$Var[u(\theta)]=E\left[\big(u(\theta)\big)^2\right]-(E[u(\theta)])^2$$

Since $E[u(\theta)]=0$, then

$$Var[u(\theta)]=E\left[\big(u(\theta)\big)^2\right]=E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right]$$

Fisher Information

The Fisher information, $I_X(\theta)$ or $I_n(\theta)$, of a random sample $X_1, X_2, \dots, X_n$ about θ is defined as

$$I_X(\theta) = Var\left[\frac{\partial}{\partial\theta}\log L(x,\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right]$$

Properties of Fisher Information:

1. $\boldsymbol{I_X(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right]}$

   Proof:

   Let $L = L(x,\theta), L' = \frac{\partial}{\partial\theta}L(x,\theta)$ and $L'' = \frac{\partial^2}{\partial\theta^2}L(x,\theta)$, then

$$\frac{\partial^2}{\partial\theta^2}\log L(x,\theta) = \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\log L(x,\theta)\right] = \frac{\partial}{\partial\theta}\left[\frac{L'}{L}\right]$$

$$= \frac{L''L-(L\prime)^2}{L^2} = \frac{L''}{L} - \frac{(L\prime)^2}{L^2}$$

$$E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right] = E\left[\frac{L''}{L} - \frac{(L\prime)^2}{L^2}\right] = E\left[\frac{L''}{L}\right] - E\left[\left(\frac{L\prime}{L}\right)^2\right]$$

   The first term in the right side can be written as

$$E\left[\frac{L''}{L}\right] = \int_{x_1} \dots \int_{x_n} \frac{L''}{L} L\, dx_1 \dots dx_n$$

$$= \frac{\partial^2}{\partial\theta^2}\int_{x_1} \dots \int_{x_n} L(x,\theta)\, dx_1 \dots dx_n = \frac{\partial^2}{\partial\theta^2}(1) = 0$$

   The second term is obtained as

$$E\left[\left(\frac{L\prime}{L}\right)^2\right] = E\left[\left(\frac{\frac{\partial}{\partial\theta}L(x,\theta)}{L(x,\theta)}\right)^2\right] = E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right]$$

   Then,

$$E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right] = E\left[\frac{L''}{L}\right] - E\left[\left(\frac{L'}{L}\right)^2\right] = 0 - E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right]$$

   This implies that,

$$I_X(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right]$$

2. $\boldsymbol{I_X(\theta) = n\, I(\theta)}$

   where $I(\theta)$ is the Fisher information at one observation defined as

$$I(\theta) = Var\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right] = E\left[\left(\frac{\partial}{\partial\theta}log f(x,\theta)\right)^2\right] = -E\left[\frac{\partial^2}{\partial\theta^2}log f(x,\theta)\right]$$

Proof:

$$I_X(\theta) = Var\left[\frac{\partial}{\partial\theta} logL(x,\theta)\right] = Var\left[\frac{\partial}{\partial\theta}\sum_{i=1}^{n}\log f(x_i;\theta)\right]$$

$$= \sum_{i=1}^{n} Var\left[\frac{\partial}{\partial\theta}\text{Log} f(x_i;\theta)\right] = n\, I(\theta)$$

3. If X and Y are two independent random samples from probability distributions $f(x,\theta)$ and $f(y,\theta)$, respectively, then

$$\boldsymbol{I_{X,Y}(\theta) = I_X(\theta) + I_Y(\theta)}$$

Proof:

$$I_{X,Y}(\theta) = E\left[\left(\frac{\partial}{\partial\theta}\log L(x,y,\theta)\right)^2\right]$$

$$= E\left[\left(\frac{\partial}{\partial\theta}\log\big(L(x,\theta)L(y,\theta)\big)\right)^2\right] \text{(Since X and Y are independent)}$$

$$= E\left[\left(\frac{\partial}{\partial\theta}\log L(x,\theta) + \frac{\partial}{\partial\theta}\log L(y,\theta)\right)^2\right]$$

$$= E\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2 + E\left(\frac{\partial}{\partial\theta}\log L(y,\theta)\right)^2$$

$$+2E\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)E\left(\frac{\partial}{\partial\theta}\log L(y,\theta)\right)$$

$$= E\left(\frac{\partial}{\partial\theta}\log L(x,\theta)\right)^2 + E\left(\frac{\partial}{\partial\theta}\log L(y,\theta)\right)^2$$

$$= I_X(\theta) + I_Y(\theta)$$

Examples 3.12:

Let $X_1, X_2, \dots, X_n$ be a random sample from normal distribution with parameters 0 and θ. Find the Fisher information of θ, $I_x(\theta)$.

Solution:

We know that the normal distribution when $\mu = 0$ and $\sigma^2 = \theta$ is given by

$$f(x,\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}, -\infty < x < \infty$$

The likelihood and the log-likelihood functions are then obtained as

$$L(x,\theta) = (2\pi\theta)^{-\frac{n}{2}}\ e^{-\frac{1}{2\theta}\sum_{i=1}^{n} {x_i}^2}$$

$$\log L(x,\theta) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\theta) - \frac{1}{2\theta}\sum_{i=1}^{n} {x_i}^2$$

1. $I_X(\theta) = Var\left[\frac{\partial}{\partial\theta}\log L(x,\theta)\right]$

   From the log-likelihood function, we get the first partial derivative with respect to θ as

$$\frac{\partial}{\partial\theta}\log L(x,\theta) = -\frac{n}{2\theta} + \frac{\sum_{i=1}^{n} {x_i}^2}{2\theta^2}$$

$$I_X(\theta) = Var\left[-\frac{n}{2\theta} + \frac{\sum_{i=1}^{n} {X_i}^2}{2\theta^2}\right] = \frac{1}{4\theta^2}Var\left[\frac{\sum_{i=1}^{n} {X_i}^2}{\theta}\right]$$

Note that $\frac{\sum_{i=1}^{n} {X_i}^2}{\theta} = \sum_{i=1}^{n} Z_i^2 \sim \chi_n^2$, then $Var\left[\frac{\sum_{i=1}^{n} {X_i}^2}{\theta}\right] = 2n$, and this implies that

$$I_X(\theta) = \frac{2n}{4\theta^2} = \frac{n}{2\theta^2}$$

2. $I_X(\theta) = nI(\theta)$

$$\log f(x,\theta) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\theta) - \frac{x^2}{2\theta}$$

$$\frac{\partial}{\partial\theta}\log f(x,\theta) = -\frac{1}{2\theta} + \frac{x^2}{2\theta^2}$$

$$I_X(\theta) = n\,I(\theta) = n\,Var\left[\frac{\partial}{\partial\theta}log f(x,\theta)\right] = n\,Var\left[-\frac{1}{2\theta} + \frac{X^2}{2\theta^2}\right]$$

$$= \frac{n}{4\theta^2}\ Var\left[\frac{X^2}{\theta}\right]$$

Since, $\frac{X^2}{\theta} = Z^2 \sim \chi_1^2$, then $Var\left[\frac{X^2}{\theta}\right] = 2$, therefore we get

$$I_X(\theta) = \frac{n}{2\theta^2}$$

3. $I_X(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right]$

   First, we should find the second partial derivative of log-likelihood function with respect to θ, which is equal to

$$\frac{\partial^2}{\partial\theta^2}\log L(x,\theta) = \frac{n}{2\theta^2} - \frac{\sum_{i=1}^{n} {x_i}^2}{\theta^3}$$

$$I_X(\theta) = -E\left[\frac{\partial^2}{\partial\theta^2}\log L(x,\theta)\right] = -E\left[\frac{n}{2\theta^2} - \frac{\sum_{i=1}^{n} {X_i}^2}{\theta^3}\right]$$

$$= -\frac{n}{2\theta^2} + \frac{\sum_{i=1}^{n} E({X_i}^2)}{\theta^3}$$

From definition of variance,

$$Var(X) = E(X^2) - \left(E(X)\right)^2 \Rightarrow E(X^2) = Var(X) + \left(E(X)\right)^2 = \theta + 0 = \theta$$

Then,

$$I_X(\theta) = -\frac{n}{2\theta^2} + \frac{n\theta}{\theta^3} = \frac{n}{2\theta^2}$$

Regularity Conditions:

(i) $\log L(x,\theta)$ or $\log f(x,\theta)$ is differentiable for all θ.

(ii) $\frac{\partial}{\partial\theta}\int_{x_1} \dots \int_{x_n} L(x;\theta)\, dx_1, \dots, dx_n = \int_{x_1} \dots \int_{x_n} \frac{\partial}{\partial\theta} L(x;\theta) dx_1, \dots, dx_n$

(iii) $\frac{\partial}{\partial\theta}\int_{x_1} \dots \int_{x_n} t(x_1, \dots, x_n)\, L(x;\theta) dx_1 \dots dx_n$

$$= \int_{x_1} \dots \int_{x_n} t(x_1, \dots, x_n)\, \frac{\partial}{\partial\theta} L(x;\theta) dx_1 \dots dx_n$$

(iv) $0 < E\left[\frac{\partial}{\partial\theta}\log L(x;\theta)\right]^2 < \infty$, for all θ.

3.2.7 Minimum Variance Unbiased Estimator (MVUE)

If a statistic T be an estimator for a parameter $\tau(\theta)$, is called to be a MVUE for $\tau(\theta)$ if

1. $E(T) = \tau(\theta)$ unbiased estimator of $\tau(\theta)$.
2. $Var(T)$ has minimum variance compared to any other unbiased estimator.

Theorem 3.10: Cramér-Rao Lower Bound (CRLB)

Let $X_1, \dots, X_n$ be a random sample from $f(x,\theta)$ and $T(X_1, \dots, X_n)$ be an unbiased estimator of $\tau(\theta)$ such that $\tau(\theta)$ is differentiable function of θ. Then, under the regularity conditions, the minimum variance of any unbiased estimator T is

$$Var(T) \geq \frac{\left(\tau'(\theta)\right)^2}{nI(\theta)}$$

Proof:

Since T is an unbiased estimator of $\tau(\theta)$ [i.e. $E(T) = \tau(\theta)$]. Then, under the regularity conditions, we get

$$\acute{\tau}(\theta) = \frac{\partial}{\partial\theta}\tau(\theta) = \frac{\partial}{\partial\theta} E(T) = \frac{\partial}{\partial\theta}\int \dots \int t(x_1, \dots, x_n) L(x;\theta)\, dx_1 \dots dx_n$$

$$= \int \dots \int t(x_1, \dots, x_n) \frac{\partial}{\partial\theta} L(x;\theta)\, dx_1 \dots dx_n$$

$$-\tau(\theta) \frac{\partial}{\partial\theta} \int \dots \int L(x;\theta)\, dx_1 \dots dx_n$$

$$= \int \dots \int t(x_1, \dots, x_n) \frac{\partial}{\partial\theta} L(x;\theta)\, dx_1 \dots dx_n$$

$$-\tau(\theta) \int \dots \int \frac{\partial}{\partial\theta} L(x;\theta)\, dx_1 \dots dx_n$$

$$= \int \dots \int [t(x_1, \dots., x_n) - \tau(\theta)] \frac{\partial}{\partial\theta} L(x;\theta)\, dx_1 \dots dx_n$$

$$= \int \dots \int [t(x_1, \dots., x_n) - \tau(\theta)] \left[\frac{\partial}{\partial\theta} \log L(x;\theta)\right] L(x;\theta) dx_1 \dots dx_n$$

$$= E\left[[t(x_1, \dots., x_n) - \tau(\theta)] \left[\frac{\partial}{\partial\theta} \log L(x;\theta)\right]\right]$$

Now by the Cauchy-Schwarz inequality

$$[\acute{\tau}(\theta)]^2 \leq E[t(x_1, \dots., x_n) - \tau(\theta)]^2\, E\left[\frac{\partial}{\partial\theta} \log L(x;\theta)\right]^2$$

or $$Var[T] \geq \frac{[\acute{\tau}(\theta)]^2}{nI(\theta)}$$

Remark: If there exists an unbiased estimator T of $\tau(\theta)$ that its variance attains the $CRLB = \frac{[\acute{\tau}(\theta)]^2}{nI(\theta)}$, then T is an MVUE estimator of $\tau(\theta)$.

3.2.8 Efficiency

An unbiased estimator T of $\tau(\theta)$ is called an **efficient** estimator of $\tau(\theta)$ if and only if

$$eff(T) = \frac{CRLB}{Var(T)} = 1$$

Theorem 3.11:

If T_1 and T_2 are both unbiased estimators of $\tau(\theta)$, then the efficiency of T_1 and T_2 is defined as follows

$$eff(T_1, T_2) = \frac{Var(T_1)}{Var(T_2)} = \begin{cases} > 1, & T_2 \text{ is more efficient than } T_1 \\ 1, & T_1 \text{ and } T_2 \text{ are equally efficient} \\ < 1, & T_1 \text{ is more efficient than } T_2 \end{cases}$$

Asymptotic Efficiency

An unbiased estimator T of $\tau(\theta)$ is called an **asymptotically efficient** estimator of $\tau(\theta)$ if

$$\lim_{n\to\infty} eff(T) = \lim_{n\to\infty} \frac{CRLB}{Var(T)} = 1$$

Example 3.13:

If $X_1, X_2, \dots, X_n$ has an exponential distribution with parameter $\frac{1}{\lambda}$. Let T_1 and T_2 are unbiased estimates of λ and $\frac{1}{\lambda}$, respectively. Find CRLB of T_1 and T_2.

Solution:

The pdf of the exponential distribution with parameter $\frac{1}{\lambda}$ is given by

$$f(x,\lambda) = \lambda e^{-\lambda x}\,, x > 0$$

Then, the likelihood and the log-likelihood functions are obtained as

$$L(x,\lambda) = \prod_{i=1}^{n} f(x_i,\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i}$$

$$\log L\,(x,\lambda) = \sum_{i=1}^{n} \log f(x_i,\lambda) = n\log\lambda - \lambda \sum_{i=1}^{n} x_i$$

Taking the first and second partial derivatives of the log-likelihood function with respect to λ, we get

$$\frac{\partial}{\partial\lambda}\log L(x,\lambda) = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i$$

$$\frac{\partial^2}{\partial\lambda^2}\log L(x,\lambda) = -\frac{n}{\lambda^2}$$

Then, the Fisher information of λ is derived as

$$I_X(\lambda) = -E\left[\frac{\partial^2}{\partial\lambda^2}\log L(x,\lambda)\right] = \frac{n}{\lambda^2}$$

Now, we want to find the CRLB for T_1 and T_2 of the two cases when $\tau(\lambda) = \lambda$ and when $\tau(\lambda) = \frac{1}{\lambda}$.

The case when $\tau(\lambda) = \lambda \Rightarrow \tau'(\lambda) = 1$, then the CRLB for T_1 is

$$CRLB(T_1) = \frac{\left(\tau'(\lambda)\right)^2}{I_X(\lambda)} = \frac{1}{n/\lambda^2} = \frac{\lambda^2}{n}$$

The second case, when $\tau(\lambda) = \frac{1}{\lambda} \Rightarrow \tau'(\lambda) = -\frac{1}{\lambda^2}$, then the CRLB for T_2 is

$$CRLB(T_2) = \frac{\left(\tau'(\lambda)\right)^2}{I_X(\lambda)} = \frac{(-1/\lambda^2)^2}{n/\lambda^2} = \frac{1}{n\lambda^2}$$

Note that $CRLB(T_2) < CRLB(T_1)$ and

$$Var(T_2) = Var(\bar{X}\ ) = \frac{\sum_{i=1}^{n} Var(X_i)}{n^2} = \frac{1}{n^2}\frac{n}{\lambda^2} = \frac{1}{n\lambda^2}$$

Then, T_2 is an efficient estimator of $\frac{1}{\lambda}$ such that

$$eff(T_2) = \frac{CRLB(T_2)}{Var(T_2)} = 1$$

Remark: T_2 is the MLE of $\frac{1}{\lambda}$.

Example 3.14:

Let $X_1, X_2, \dots, X_n \sim Poisson(\lambda)$. Find CRLB of the MLE of λ and prove it is an efficient estimator.

Solution:

From Example 3.1 and Example 3.4, we get the MLE of λ is $T = \bar{X}$ and it is an unbiased estimator of λ where

$$\tau(\lambda) = \lambda \Rightarrow \tau'(\lambda) = 1$$

The pdf of Poisson distribution with parameter λ is defined as

$$f(x,\lambda) = \frac{e^{-\lambda}\,\lambda^x}{x!}, x = 0, 1, 2, \dots$$

The logarithm function of the pdf and the derivatives are

$$\log f(x,\lambda) = x \log\lambda - \lambda - \log x!$$

$$\frac{\partial \log f(x,\lambda)}{\partial\lambda} = \frac{x}{\lambda} - 1$$

$$\frac{\partial^2 \log f(x,\lambda)}{\partial\lambda^2} = -\frac{x}{\lambda^2}$$

Then, the Fisher information is given as

$$I_X(\lambda) = nI(\lambda) = -nE\left[\frac{\partial^2 \log f(x,\lambda)}{\partial \lambda^2}\right] = nE\left(\frac{X}{\lambda^2}\right) = \frac{n}{\lambda}$$

where $E(X) = \lambda$. Therefore, the CRLB is equal to

$$CRLB = \frac{\left(\tau'(\lambda)\right)^2}{I_X(\lambda)} = \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n}$$

Note that $Var(\bar{X}) = \frac{\lambda}{n}$ and thus the variance of the MLE equals to the CRLB. Therefore, the MLE, $\bar{X}$, is an efficient estimator of λ.

Example 3.15:

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Show that,

(i) $\bar{X}$ is an efficient estimator of μ.

(ii) $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an asymptotically efficient of σ^2.

(iii) $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is an asymptotically efficient of σ^2.

(iv) $S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$ is an efficient estimator of σ^2 if μ is known.

Solution :

The pdf of the $N(\mu, \sigma^2)$, the likelihood and the log-likelihood functions are

$$f(x, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, x > 0$$

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

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

The first and second partial derivatives with respect to μ and σ^2 are

$$\frac{\partial \log L}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu), \qquad \frac{\partial \log L}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$\frac{\partial^2 \log L}{\partial \mu^2} = -\frac{n}{\sigma^2}, \qquad \frac{\partial^2 \log L}{\partial \sigma^4} = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{n}(x_i - \mu)^2$$

Now, to study the efficiency, we need to determine the unbiasedness, CRLB and the variance:

(i) The efficiency of $\bar{X}$:

From Example 3.6, we get

$$E(\bar{X}) = \mu \text{ and } Var(\bar{X}\ ) = \frac{\sigma^2}{n}$$

i.e. $\bar{X}$ is an unbiased estimator of μ. Now, the Fisher information of μ is given as

$$I_X(\mu) = -E\left[\frac{\partial^2}{\partial\mu^2}\log L(\mu,\sigma^2)\right] = \frac{n}{\sigma^2}$$

Thus, the CRLB of $\bar{X}$ is

$$CRLB(\bar{X}) = \frac{\left(\tau'(\mu)\right)^2}{I_X(\mu)} = \frac{\sigma^2}{n}$$

which is equal to the variance of $\bar{X}$, then we conclude that the estimator $\bar{X}$ is an efficient of μ. Notice that, $\bar{X}$ is the MLE of μ.

(ii) The efficiency of $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$:

$$E(\ S^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$$

We know from Section 2.3, when $X_i \sim N(\mu,\sigma^2), i = 1,2,..,n$, then

$$\frac{(n-1)S^2}{\sigma^2} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim X_{n-1}^2$$

and $E(S^2) = \sigma^2$, $Var(S^2) = \frac{2\sigma^4}{n-1}$. Thus, S^2 is an unbiased estimator of σ^2. The Fisher information of σ^2 is given by

$$I_X(\sigma^2) = -E\left[\frac{\partial^2}{\partial\sigma^4}\log L(\mu,\sigma^2)\right] = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6}\sum_{i=1}^{n}E(X_i - \mu)^2$$

From Corollary 2.2, when $X_i \sim N(\mu,\sigma^2), i = 1,2,..,n$, then

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2 \quad \text{and} \quad E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = n.$$

Therefore,

$$I_X(\sigma^2) = -\frac{n}{2\sigma^4} + \frac{n}{\sigma^4} = \frac{n}{2\sigma^4}$$

Now, the CRLB is obtained as

$$CRLB(\ S^2) = \frac{\left(\tau'(\sigma^2)\right)^2}{I_X(\sigma^2)} = \frac{1}{n/2\sigma^4} = \frac{2\sigma^4}{n}$$

$$eff(S^2) = \frac{CRLB(S^2)}{Var(S^2)} = \frac{2\sigma^4/n}{2\sigma^4/n-1} = \frac{n-1}{n}$$

$$\lim_{n\to\infty} eff\,(S^2) = \lim_{n\to\infty} \frac{n-1}{n} = \lim_{n\to\infty}\left(1-\frac{1}{n}\right) = 1$$

Then, S^2 is asymptotically efficient of σ^2.

(iii) The efficiency of $S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$:

From Example 3.6,

$$E(S_1^2) = \frac{(n-1)\sigma^2}{n} \text{ and } Var(S_1^2) = \frac{2(n-1)\sigma^4}{n^2}$$

i.e. S_1^2 is not an unbiased estimator of σ^2. The CRLB of S_1^2 is obtained as

$$CRLB(S_1^2) = \frac{\left(\tau'(\sigma^2)\right)^2}{I_X(\sigma^2)} = \frac{1}{n/2\sigma^4} = \frac{2\sigma^4}{n}$$

$$eff\,(S_1^2) = \frac{CRLB(S_1^2)}{Var(S_1^2)} = \frac{2\sigma^4/n}{2(n-1)\sigma^4/n^2} = \frac{n}{n-1}$$

$$\lim_{n\to\infty} eff\,(S_1^2) = \lim_{n\to\infty}\frac{n}{n-1} = \lim_{n\to\infty}\frac{1}{1-\frac{1}{n}} = 1$$

Then S_1^2 is an asymptotically efficient of σ^2.

(iv) The efficiency of $S_2^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2$:

From Corollary 2.2, when $X_i \sim N(\mu,\sigma^2), i = 1,2,..,n$, then

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2 \text{ and } E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = n \text{ and } Var\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = 2n$$

Therefore, the mean and the variance of S_2^2 are calculated as follows

$$E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = \frac{n}{\sigma^2}E\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}\right] = n \Rightarrow E(S_2^2) = \sigma^2$$

$$Var\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right] = \frac{n^2}{\sigma^4}Var\left[\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{n}\right] = 2n \Longrightarrow Var(S_2^2) = \frac{2\sigma^4}{n}$$

Now, the CRLD and the efficiency of S_2^2 are

$$CRLB(S_2^2) = \frac{\left(\tau'(\sigma^2)\right)^2}{I_X(\sigma^2)} = \frac{1}{n/2\sigma^4} = \frac{2\sigma^4}{n}$$

$$eff(S_2^2) = \frac{CRLB(\,S_2^2)}{Var(\,S_2^2)} = \frac{2\sigma^4/n}{2\sigma^4/n} = 1$$

Thus, S_2^2 is an efficient estimator of σ^2.

When the CRLB Doesn't Help:

The Cramer-Rao lower bound gives a necessary and sufficient condition for the existence of an efficient estimator. However, MVUEs are not necessarily efficient. What can we do in such cases? The Rao-Blackwell theorem, when applied in combination with a complete sufficient statistic, gives another way to find MVUEs that applies even when the CRLB is not defined.

Theorem 3.12: (Rao-Blackwell Theorem)

Let $X_1, \dots, X_n$ be a random sample from $f(x,\theta), \theta$ may be a vector of parameters; and let $S_1 = s_1(X_1, \dots, X_n), \dots, S_k = s_k(X_1, \dots, X_n)$ be a set of jointly sufficient statistics. Let the statistic $T = t(X_1, \dots, X_n)$ be an unbiased estimator of $\tau(\theta)$. Define,

$$T' = E(T|S_1, \dots, S_k)$$

Then,

1. T' is a statistic and it is a function of the sufficient statistics $S_1, \dots, S_k$. Write
$$T' = t'(S_1, \dots, S_k)$$
2. T' is an unbiased estimator of $\tau(\theta)$; $E(T') = \tau(\theta)$.
3. $Var(T') \le Var(T)$ for all θ, and $Var(T') = Var(T)$ iff $T' = T$.

Proof:

1. $S_1, \dots., S_k$ are sufficient statistics; so, the conditional distribution of any statistic T, given $S_1, \dots., S_k$ is independent of θ, hence $T' = E[T|S_1, \dots., S_k]$ is independent of θ, and so T' is a statistic which is obviously a function of $S_1, \dots., S_k$.

2. $E[T'] = E\big[E[T|S_1, \dots., S_k]\big] = E[T] = \tau(\theta)$ [using $E[Y] = E\big[E[Y|X]\big]$].

3. we can write

$$MSE[T] = Var[T] = E[(T - E[T'])^2] = E[(T - T' + T' - E[T'])^2]$$
$$= E[(T - T')^2] + 2E[(T - T')(T' - E[T'])] + E[(T' - E[T'])^2]$$

$$= E[(T-T')^2] + 2E[(T-T')(T'-E[T'])] + Var[T']$$

But

$$E[(T-T')(T'-E[T'])] = E\big[E[(T-T')(T'-E[T'])|S_1, \dots, S_k]\big]$$

and

$E[(T-T')(T'-E[T'])|S_1 = s_1; \dots; S_k = s_k]$

$$= \{t'(s_1, \dots, s_k) - E[T']\}E[(T-T')|S_1 = s_1; \dots; S_k = s_k]$$

$$= \{t'(s_1, \dots, s_k) - E[T']\}(E[T|S_1 = s_1; \dots; S_k = s_k] - E[T'|S_1 = s_1; \dots; S_k = s_k])$$

$$= \{t'(s_1, \dots, s_k) - E[T']\}[t'(s_1, \dots, s_k) - t'(s_1, \dots, s_k)] = 0$$

and therefore

$$Var[T] = E[(T-T')^2] + Var[T'] \geq Var[T']$$

Note that $Var[T] > Var[T']$ unless T equals T' with probability 1.

Example 3.16:

Let $X_1, \dots, X_n$ be a random sample from the Bernoulli(p)

$$f(x;p) = p^x q^{1-x}, \; x = 0 \text{ or } 1$$

and let $T = X_1$ be an unbiased estimate of θ. Find a MVUE of p.

Solution:

Since, $T = X_1$ is an unbiased estimator such that $E(T) = E(X_1) = p$. From Example 3.9, we get $S = \sum_{i=1}^{n} X_i$ is a sufficient statistic. According to the Rao-Blackwell Theorem

$T' = E(T|S) = E(X_1 | \sum_{i=1}^{n} X_i) = \sum_{x_1=0}^{1} X_1 \, P(X_1 | \sum_{i=1}^{n} X_i)$

$$= (0)P(X_1 = 0|\textstyle\sum_{i=1}^{n} X_i = S) + (1)P(X_1 = 1|\textstyle\sum_{i=1}^{n} X_i = S)$$

$$= \frac{P\left(X_1 = 1, \sum_{i=1}^{n} X_i = S\right)}{P\left(\sum_{i=1}^{n} X_i = S\right)}$$

$$= \frac{P(X_1 = 1)P\left(\sum_{i=2}^{n} X_i = S-1\right)}{P\left(\sum_{i=1}^{n} X_i = S\right)}$$

$$= \frac{p\binom{n-1}{S-1} p^{S-1} q^{n-S}}{\binom{n}{S} p^S q^{n-S}}$$

$$= \frac{(n-1)!}{(S-1)!(n-S)!} \frac{S!(n-S)!}{n!}$$

$$= \frac{S}{n} = \bar{X}$$

Thus, $T' = \bar{X}$ is a statistic and a function of a sufficient statistic S and an unbiased estimator of p where $E(T') = E(\bar{X}) = p$. Therefore, $T' = \bar{X}$ is a MVUE of p with minimum variance such that

$$Var(T') = Var(\bar{X}) = Var\left(\frac{\sum_{i=1}^{n} X_i}{n}\right) = \frac{1}{n^2} npq = \frac{pq}{n}$$

While, $$V(T) = V(X_1) = pq$$

Thus, $$V(T') < V(T)$$

Theorem 3.13: (Lehman-Scheffé Theorem)

Let $X_1, \ldots, X_n$ be a random sample from $f(x, \theta)$, θ may be a vector of parameters $(\theta_1, \ldots, \theta_k)$. If $S = s(S_1, \ldots, S_k)$ is a complete sufficient statistic and if $T^* = t^*(S)$ a function of S, is an unbiased estimator of $\tau(\theta)$. Then, $T^*$is UMVUE of $\tau(\theta)$.

Proof:

Let T' be any unbiased estimator of $\tau(\theta)$ which is a function of S; that is, $T' = t'(S)$. Then $E[T^* - T'] = 0$ for all $\theta \in \emptyset$, and $T^* - T'$ is a function of S; so by completeness of $S, P[t^*(S) = t'(S)] = 1$ for all $\theta \in \emptyset$. Hence there is only one unbiased estimator of $\tau(\theta)$ that is function of S. Now let T be any unbiased estimator of $\tau(\theta)$. T^* must be equal to $E[T|S]$ since $E[T|S]$ is an unbiased estimator of $\tau(\theta)$ depending on S. By Theorem 3.11, $Var[T^*] \leq Var[T]$ for all $\theta \in \emptyset$; so $T^*$is an UMVUE.

Example 3.17:

Let $X_1, X_2, \ldots, X_n$ be a random sample from the $Exponential(\beta)$,

$$f(x, \beta) = \frac{1}{\beta} e^{-x/\beta}, \;\; x > 0$$

Find UMVUE of β and $\frac{1}{\beta}$.

Solution:

Since the exponential distribution is a member of the exponential family, then $S = \sum_{i=1}^{n} X_i$ is a complete sufficient statistic. Thus, we need to derive two functions of S that are unbiased estimators of β and $\frac{1}{\beta}$.

1. Put $T_1^* = cS$, c is a constant such that
$$E(T_1^*) = \beta \Rightarrow E(cS) = \beta \Rightarrow c\, E(\textstyle\sum_{i=1}^{n} X_i) = \beta \Rightarrow c\, n\beta = \beta \Rightarrow c = \frac{1}{n}$$
Thus, $T_1^* = cS = \bar{X}$ is a UMVUE of β.
2. Put $T_2^* = \frac{c}{S}$, c is a constant such that
$$E(T_2^*) = E\left(\frac{c}{S}\right) = c\, E\left(\frac{1}{\sum_{i=1}^{n} X_i}\right) = \frac{1}{\beta}$$
Since, $S = \sum_{i=1}^{n} X_i \sim Gamma\ (n, \beta)$, then
$$E\left(\frac{1}{S}\right) = \int_0^\infty \frac{1}{s} \frac{1}{\Gamma(n)\beta^n} s^{n-1} e^{-s/\beta}\, ds = \frac{\Gamma(n-1)\beta^{n-1}}{\Gamma(n)\beta^n} = \frac{1}{(n-1)\beta}$$
Thus,
$$E(T_2^*) = c\, E\left(\frac{1}{S}\right) = c\, \frac{1}{(n-1)\beta} = \frac{1}{\beta} \Rightarrow c = n-1$$
Therefore, $T_2^* = \frac{c}{S} = \frac{n-1}{\sum_{i=1}^{n} X_i}$ is a UMVUE of $\frac{1}{\beta}$.

3.3 Properties of Maximum Likelihood Estimators

Let $X_1, X_2, \dots, X_n$ be a random sample with probability distribution $f(x, \theta)$. If $MLE = \hat{\theta}$ of θ and under certain regularity conditions, then $\hat{\theta}$ satisfies the following properties:

1. **Invariance:** Let $h(\theta)$ be a function of θ. Then, $T = h(\hat{\theta})$ is the MLE of $h(\theta)$.
2. **Sufficiency:** If a sufficient statistic exists for θ, the MLE of θ must be a function of it.
3. **Asymptotically unbiased:** $\lim_{n\to\infty} E(\hat{\theta}) = \theta$
4. **Consistency:** $\lim_{n\to\infty} P\left(\left|\hat{\theta} - \theta\right| \geq \varepsilon\right) = 0\,,\ \forall\theta$
5. **Asymptotic efficiency:** If a most efficient unbiased estimator T of θ exists (i.e. T is unbiased and its variance is equal to the CRLB). Then, the maximum likelihood method of estimation will produce it.

6. **Asymptotic normality:** The MLE $\hat{\theta}$ of θ has asymptotic normal distribution such that
$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right), n \to \infty \text{ where } Var(\hat{\theta}) = CRLB(\hat{\theta}) = \frac{1}{nI(\theta)}.$$
In general, if $\hat{\tau}(\theta)$ be the MLE of $\tau(\theta)$, then $\hat{\tau}(\theta)$ has distribution as
$$\sqrt{n}(\hat{\tau}(\theta)-\tau(\theta)) \xrightarrow{d} N\left(0, \frac{(\tau\prime(\theta))^2}{I(\theta)}\right) \text{ or } \hat{\tau}(\theta) \xrightarrow{d} N\left(\tau(\theta), \frac{(\tau\prime(\theta))^2}{nI(\theta)}\right).$$

3.4 Location and Scale Invariance

3.4.1 Location Invariance:

Location Parameter:

Let $f(x)$ be any pdf. The family of pdfs $f(x-\mu)$ indexed by parameter μ is called the **location family** with standard pdf $f(x)$ and μ is the **location parameter** for the family.

Equivalently, μ is a location parameter for $f(x)$ iff the distribution $f(x-\mu)$ does not depend on μ.

Location Invariant:

Let $X_1, X_2, \dots, X_n$ be a random sample of a distribution with pdf (or pmf); $f(x, \mu)$; $\mu \in \Omega$.

- An estimator $t(x_1, \dots, x_n)$ is defined to be a **location equivariant** iff
$$t(x_1+c, \dots, x_n+c) = t(x_1, \dots, x_n) + c \text{ for all values } c.$$
- An estimator $t(x_1, \dots, x_n)$ is defined to be a **location invariant** iff
$$t(x_1+c, \dots, x_n+c) = t(x_1, \dots, x_n) \text{ for all values } c.$$

Example 3.18:

- If $X \sim N(\theta, 1)$, then the distribution of $X-\theta \sim N(0,1)$ is independent of θ $\longrightarrow$ θ is a location parameter.
- Let $t(x_1, \dots, x_n) = \bar{X}$. Then,

$$t(x_1+c, \dots, x_n+c) = \frac{x_1+c+\dots+x_n+c}{n} = \frac{x_1+\dots+x_n+nc}{n}$$

$$= \bar{X} + c = t(x_1, \dots, x_n) + c$$

$\longrightarrow$ $\bar{X}$ is location equivariant.

- Let $t(x_1, \dots, x_n) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Then,

$$t(x_1 + c, \dots, x_n + c) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i + c - (\bar{X} + c)\right)^2$$

$$= S^2 = t(x_1, \dots, x_n)$$

$\longrightarrow$ S^2 location invariant.

3.4.2 Scale Invariant:

Scale Parameter:

Let $f(x)$ be any pdf. The family of pdfs $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$ for $\sigma > 0$, indexed by parameter σ is called the **scale family** with standard pdf $f(x)$ and σ is the **scale parameter** for the family.

Equivalently, σ is a scale parameter for $f(x)$ iff the distribution $\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$ does not depend on σ.

Scale Invariant:

Let $X_1, X_2, \dots, X_n$ be a random sample of a distribution with pdf (or pmf); $f(x, \sigma)$; $\sigma \in \Omega$.

- An estimator $t(x_1, \dots, x_n)$ is defined to be a **scale equivariant** iff

$$t(c\, x_1, \dots, c\, x_n) = c\; t(x_1, \dots, x_n) \text{ for all values } c.$$

- An estimator $t(x_1, \dots, x_n)$ is defined to be a **scale invariant** iff

$$t(c\, x_1, \dots, c\, x_n) = t(x_1, \dots, x_n) \text{ for all values } c.$$

Example 3.19:

- If $X \sim Exponential\left(\frac{1}{\theta}\right)$, then the distribution $\frac{1}{\theta} f\left(\frac{x}{\theta}\right)$ is independent of θ $\longrightarrow$ θ is a scale parameter.

- Let $t(x_1, \dots, x_n) = \bar{X}$. Then,

$$t(c\,x_1,\dots,c\,x_n) = \frac{c\,(x_1+\cdots+x_n)}{n} = c\,\bar{X} = c\,t(x_1,\dots,x_n)$$

$\longrightarrow\ \bar{X}$ is scale equivariant.

- Let $t(x_1,\dots,x_n) = \frac{X_1}{X_1+X_2}$. Then,

$$t(c\,x_1,\dots,c\,x_n) = \frac{c\,X_1}{c\,X_1+c\,X_2} = \frac{X_1}{X_1+X_2} = t(x_1,\dots,x_n)$$

$\longrightarrow\ \frac{X_1}{X_1+X_2}$ is scale invariant.

3.4.3 Location-Scale Invariant:

Location-Scale Parameter:

Let $f(x)$ be any pdf. The family of pdfs $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$ for $\sigma>0$, indexed by parameter (μ,σ) is called the **location-scale family** with standard pdf $f(x)$ and μ is a **location parameter** and σ is the **scale parameter** for the family.

Equivalently, μ is a location parameter and σ is a scale parameter for $f(x)$ iff the distribution $\frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$ does not depend on μ and σ.

Location-Scale Invariant:

Let $X_1, X_2, \dots, X_n$ be a random sample of a distribution with pdf (or pmf); $f(x,\sigma)$; $\sigma\in\Omega$.

- An estimator $t(x_1,\dots,x_n)$ is defined to be a **location-scale equivariant** iff

  $t(c\,x_1+d,\dots,c\,x_n+d) = c\,t(x_1,\dots,x_n)+d$ for all values $c>0$ and d.
- An estimator $t(x_1,\dots,x_n)$ is defined to be a **location-scale invariant** iff
  $t(c\,x_1+d,\dots,c\,x_n+d) = t(x_1,\dots,x_n)$ for all values $c>0$ and d.

Example 3.20:

- If $X\sim N(\mu.\sigma^2)$, then the distribution of $Y=\frac{X-\mu}{\sigma}\sim N(0,1)$ is independent of μ and $\sigma^2 \longrightarrow \mu$ and σ^2 are location-scale parameters.

- Let $t(x_1,\dots,x_n)=\bar{X}$. Then,

$$t(cx_1 + d, \dots, cx_n + d) = \frac{c(x_1 + \cdots + x_n) + nd}{n} = c\bar{X} + d = c\, t(x_1, \dots, x_n) + d$$

$\longrightarrow$ $\bar{X}$ is location-scale equivariant.

- Let $t(x_1, \dots, x_n) = \frac{Y_n - Y_1}{S}$. Then,

$$t(c\, x_1 + d, \dots, c\, x_n + d) = \frac{(cY_n + d) - (cY_1 + d)}{cS + d}$$

$$= \frac{cY_n - cY_1}{cS} = \frac{Y_n - Y_1}{S} = t(x_1, \dots, x_n)$$

$\longrightarrow$ $\frac{Y_n - Y_1}{S}$ is location-scale invariant.