King Saud University
College of Science
Department of Statistics and Operations Research

STAT 223
Theory of Statistics 1

Lectures' Notes

1438/1439

Dr. Samah Alghamdi

Table of Contents

Outline of the Course:

Course code and number: STAT 223.

Course name: Theory of Statistics 1.

Credits: 3(2+1).

Pre-requisite: STAT 215.

Instructor:

Dr. Samah Abdullah Alghamdi.

Office: 346 – Building 5.

E-mail : samalghamdi@ksu.edu.sa

Web Site: http://fac.ksu.edu.sa/samalghamdi

Text Book:

Probability and Statistics for Engineers and Scientists, 2012, Ninth Edition by R. Walpole, R. Myers, S. Myers and K. Ye, Person.

References:

1. Introduction to Mathematical Statistics, 2005, Sixth Edition by R. Hogg, J. McKean, and A. Craig, Prentice Hall.
2. Introduction to the Theory of Statistics, 2007, Third Edition by A. Mood, F. Graybill and D. Boes, McGrow-Hill.
3. Statistical Inference, 2002, Second Edition by G. Casella and R. Berger, Duxbury.

Main Topics:

1. Sampling Distribution and Central Limit Theorem.
2. Point Estimation.
3. Estimator Properties: Unbiasedness – Mean Squared Error – Consistency – Sufficiency – Minimal Sufficiency – Completeness - Minimum Variance Unbiased Estimator.
4. Exponential Family Theorem.

5. Fisher Information and Cramer-Rao Inequality.
6. Rao-Blackwell and Lehmann-Scheffé Theorems.
7. Estimation Methods: Method of Moments – Maximum Likelihood Method.
8. Interval Estimation: Pivotal Quantity – Large Sample Confidence Interval.
9. Bayesian Estimate: Prior and Posterior Distributions – Loss Function Approach – Bayesian Confidence Intervals.

Marking Scheme:

25 Marks: First Mid-Term Exam.

25 Marks: Second Mid-Term Exam.

10 Marks: Assignments and Quizzes.

40 Marks: Final Exam.

Chapter 1: Introduction

This chapter introduce a brief review of some basic definitions and statistical distributions.

1.1 Definition and Basic Concept

In this chapter, we give some basic definitions and concepts.

Population:

- A population is the largest collection of elements or individuals in which we are interested in a particular time and about which we want to make some statement or conclusion.
- The population values usually denoted by $\underline{X} = (X_1, X_2, \dots, X_N)$, where N is the number of elements in the population, called the population size.

Sample:

- A sample is a subset of a population on which we collect data.
- The sample values usually denoted by $\underline{x} = (x_1, x_2, \dots, x_n)$, where n is the number of elements in the sample, called the sample size.

Parameter:

- A parameter is a measure (or number) obtained from the population values.
- Values of the parameters are unknown in general.

Statistic:

- A statistic is a measure (or number) obtained from the sample values.
- Values of the statistic are known in general.

Random Variable:

- A random variable X is a function that associates a real number with each element in the sample space.
- Most of the time, statisticians deal with two special kinds of random variables, that are discrete and continuous random variables.

Discrete Random Variable:

A random variable X is discrete if:

1. It can take on values from finite or countable values.
2. It has a discrete distribution, called the **probability mass function** (**pmf**) of X if, for each possible outcome x

$$f_X(x) \geq 0, \quad \sum_x f_X(x) = 1, \quad \text{and} \quad f_X(x) = P(X = x).$$

Continuous Random Variable:

A random variable X is continuous if:

1. It can take on values from an interval or not countable values.
2. It has a continuous distribution, called the **probability density function** (**pdf**) for X, defined over the set of real numbers, if

$$f_X(x) \geq 0 \ \text{ for all } x \in R, \quad \int_{-\infty}^{\infty} f_X(x)dx = 1, \quad \text{and} \quad P(a \leq X \leq b) = \int_a^b f_X(x)dx \ .$$

Cumulative Distribution Function:

Let X be a random variable. The cumulative distribution function (**distribution function** or **cdf**) of X is a function such that

$$F_X(x) = P(X \leq x), \text{ for } -\infty < x < \infty.$$

Random Sample:

A random sample is a sample that is chosen randomly. Random sample are used to avoid bias and other unwanted effects.

Joint Probability distribution:

The function $f(x, y)$ is a joint probability distribution of the random variables X and Y if:

1. $f(x, y) \geq 0$, for all (x, y).
2. $\sum_x \sum_y f(x, y) = 1$ if X and Y are discrete

   $\int_x \int_y f(x, y) dy dx = 1$ if X and Y are continuous.

Independent Random Variables:

Let $X_1, X_2, \dots, X_n$ be a n random variables, discrete or continuous, with joint probability distribution $f(x_1, x_2, \dots, x_n)$. The random variables $X_1, X_2, \dots, X_n$ are said to be **mutually statistically independent** if and only if

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n).$$

For all $(x_1, x_2, \dots, x_n)$ within their range.

Expectations and Moments:

The ***r*th moment about the origin** of the random variable X is given by

$$\mu_r' = E(X^r) = \begin{cases} \sum x^r f_X(x), & \text{If } X \text{ is discrete,} \\ \int x^r f_X(x) dx, & \text{If } X \text{ is continuous.} \end{cases}$$

The first moment (**mean** or **expected value)** and the **second moment** are given by $\acute{\mu}_1 = \mu = E(X)$ and $\acute{\mu}_2 = E(X^2)$, respectively.

The **variance** is defined as

$$Var(X) = \sigma^2 = \acute{\mu}_2 - \acute{\mu}_1^2 = E[(X-\mu)^2] = E(X^2) - \big(E(X)\big)^2.$$

The standard deviation is the square root of the variance denoted as

$$\sigma = \sqrt{\sigma^2} = \sqrt{E(X^2) - \big(E(X)\big)^2}.$$

The ***r*****th central moment** of X is defined as

$$E[(X-\mu)^r] = \begin{cases} \sum (x-\mu)^r \, f_X(x), & \text{If } X \text{ is discrete,} \\ \int (x-\mu)^r \, f_X(x)dx, & \text{If } X \text{ is continuous.} \end{cases}$$

Remark:

If $Y = aX \pm b$, then the mean and the variance of Y are given by

$$E(Y) = aE(X) \pm b \quad \text{and} \quad Var(Y) = a^2 Var(X)$$

Example1.2:

Consider the following distribution:

X	-2	0	1	2	Total
$f(x)$	0.33	0.24	0.15	0.28	1

Find: 1. $P(X=0), P(X<1), P(X\geq 2)$.

2. The cdf of X.

3. The mean and the variance.

Solution:

1. $P(X=0)=0.24,\ \ P(X<1)=0.57,\ \ P(X\geq 2)=0.28$.
2. The cdf of X is

X	-2	0	1	2
$F(x)$	0.33	0.57	0.72	1

3. The mean and the variance are given as

X	-2	0	1	2	Total
$f(x)$	0.33	0.24	0.15	0.28	1
$E(X)$	-0.66	0	0.15	0.56	0.05
$E(X^2)$	1.32	0	0.15	1.12	2.59

Then,

$E(X)=0.05$ and $Var(X)=E(X^2)-\left(E(X)\right)^2=2.59-0.05^2=2.5875.$

Example 1.3:

Let X be a continuous random variable whose probability density function is

$$f(x) = 3x^2, \quad \text{for } 0 < x < 1.$$

Find:

1. Prove $f(x)$ is a pdf.
2. $P(0.5 < X < 1)$.
3. The cdf of X.
4. $E(X)$ and $Var(X)$.

Solution:

1. Since $f(x) \geq 0$ for all $x \in (0,1)$ and

$\int_0^1 f(x)dx = \int_0^1 3x^2 dx = \; x^3]_0^1 = 1$. Thus, $f(x)$ is a pdf.

2. $P(0.5 < X < 1) = \int_{0.5}^1 3x^2 dx = \; x^3]_{0.5}^1 = 1 - 0.5^3 = 0.875.$
3. $F(x) = \int_0^x 3x^2 dx = \; x^3]_0^x = x^3.$
4. $E(X) = \int_0^1 3x^3 dx = \frac{3}{4}x^4]_0^1 = \frac{3}{4}.$

   $$E(X^2) = \int_0^1 3x^4 dx = \frac{3}{5}x^5]_0^1 = \frac{3}{5}.$$

   $$Var(X) = E(X^2) - \left(E(X)\right)^2 = \frac{3}{5} - \left(\frac{3}{4}\right)^2 = 0.0375.$$

Moment-Generation Function:

The moment-generation function (**mgf**) of a random variable X is given by $E(e^{tX})$ and is denoted by, $M_X(t)$. Hence,

$$M_X(t) = E(e^{tX}) = \begin{cases} \sum_x e^{tx} f(x), & \text{if } X \text{ is discrete,} \\ \int_x e^{tx} f(x)dx, & \text{if } X \; is \text{ continuous} \end{cases}.$$

Some properties of the mgf:

1. $M_{X+a}(t) = e^{at} M_X(t)$.
2. $M_{aX}(t) = M_X(at)$.

1.2 Discrete Probability Distributions

In this section, we present some commonly used distributions for the discrete random variable.

1.2.1 Bernoulli and Binomial Distribution

A Bernoulli trial can result in a success with probability p and a failure with probability $q = 1 - p$. Then the probability of the binomial random variable X, the number of successes in n independent trials, is

$$f(x; n, p) = \binom{n}{x} p^x q^{n-x}, \;\; x = 0, 1, 2, \dots, n.$$

where $\binom{n}{x} = \frac{n!}{x!(n-x)!}$.

The mean, variance and mgf of the binomial distribution, $Binomial(n, p)$, are

$$\mu = np, \;\; \sigma^2 = npq \;\; \text{and } M(t) = (pe^t + q)^n.$$

Example 1.4:

The probability that a certain kind of component will survive a shock test is 0.75. Find:

1. The probability that exactly 2 of the next 4 components tested survive.
2. The probability that more than 2 of the next 4 components tested survive.
3. The mean and the standard deviation.

Solution:

Assuming that the tests are independent and $p = 0.75$ for each of the $n = 4$ tests, we obtain:

$$f(x; 4,0.75) = \binom{4}{x}(0.75)^x(0.25)^{4-x}, \quad x = 0, 1, 2, 3, 4.$$

1. $f(2; 4,0.75) = \binom{4}{2}0.75^2 0.25^2 = 0.2109$
2. $P(X > 2) = f(3; 4,0.75) + f(4; 4,0.75) = \binom{4}{3}0.75^3 0.25^1 + \binom{4}{4}0.75^4 0.25^0 = 0.7383$
3. $\mu = np = (4)(0.75) = 3$ and $\sigma = \sqrt{npq} = \sqrt{4(0.75)(0.25)} = 0.866.$

1.2.2 Poisson Distribution

The probability distribution of the Poisson random variable X with parameter λ, $Poisson(\lambda)$, representing the number of outcomes occurring in a given time interval or specified region denoted by *t*, is

$$f(x; \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

where $\lambda > 0$ is the average number of outcomes per unit time, distance or area.

The mean and the variance of the Poisson distribution are

$$\mu = \sigma^2 = \lambda t.$$

Example 1.5:

Births in a hospital occur randomly at an average rate of 1.6 births per hour. Calculate:

1. The probability of observing 4 births in a given hour.
2. The probability of observing more than or equal to 2 births in a given hour.
3. The mean of births per hour.
4. The probability of observing 1 birth per 2 hours.
5. The variance of births per 30 minutes.

Solution:

Let X be the number of births in a given hour and $\lambda t = 1.6$ per hour. The pdf of X is given as

$$f(x; 1.6) = \frac{e^{-1.6}(1.6)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

1. $f(4; 1.6) = \frac{e^{-1.6}(1.6)^4}{4!} = 0.0551$
2. $P(X \geq 2) = 1 - P(X < 2) = 1 - [f(1; 1.6) + f(0; 1.6)] = 0.4751$
3. $\mu = \lambda t = 1.6$
4. $\lambda t = (1.6)(2) = 3.2 \ \Rightarrow f(1; 3.2) = \frac{e^{-3.2}(3.2)^1}{1!} = 0.1304$
5. $\sigma^2 = \lambda t = (1.6)(0.5) = 0.8.$

1.3 Continuous Probability Distributions

1.3.1 Uniform Distribution

The density function of the continuous uniform random variable X on the interval $[a, b]$ is

$$f(x; a; b) = \frac{1}{b-a}, \quad a \le x \le b.$$

The mean and the variance of the uniform distribution, $Uniform(a, b)$, are

$$\mu = \frac{a+b}{2} \quad \text{and} \quad \sigma^2 = \frac{(b-a)^2}{12}.$$

Example 1.6:

Suppose that a large conference room at a certain company can be reserved for no more than 4 hours. In fact, it can be assumed that the length X of a conference has a uniform distribution on interval $[0, 4]$.

(a) What is the probability density function?
(b) What is the probability that any given conference lasts at least 3 hours?

Solution:

(a) The appropriate density function for the uniformly distributed random variable X in the situation is

$$f(x) = \frac{1}{4}, \quad 0 \le x \le 4$$

(b) $P[X \ge 3] = \int_3^4 \frac{1}{4}\, dx = \frac{1}{4}.$

1.3.2 Exponential Distribution

The pdf of the exponential distribution for a continuous random variable X with parameter $\theta > 0$, denoted as $Expnential\left(\frac{1}{\theta}\right)$, is given as

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

The mean and the variance of this distribution are

$$E(X) = \frac{1}{\theta} \quad \text{and} \quad V(X) = \frac{1}{\theta^2}.$$

The cdf and mgf obtained as

$$F(x) = 1 - e^{-\theta x} \text{ and } M(t) = \frac{\theta}{\theta - t} = \left(1 - \frac{t}{\theta}\right)^{-1}, t < \theta.$$

1.3.3 Gamma Distribution

The continuous random variable X has a gamma distribution with parameters α and $\frac{1}{\beta}$, $Gamma\left(\alpha, \frac{1}{\beta}\right)$ if its density function is given by

$$f\left(x;\alpha,\frac{1}{\beta}\right) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0$$

where $\alpha > 0, \beta > 0$ and $\Gamma(\alpha)$ is a gamma function defined as

$$\Gamma(\alpha) = (\alpha - 1)! = \int_0^\infty y^{\alpha-1} e^{-y}\, dy$$

The mean, the variance and the mgf are

$$E(X) = \frac{\alpha}{\beta}, \quad V(x) = \frac{\alpha}{\beta^2} \text{ and } M(t) = \left(\frac{\beta}{\beta - t}\right)^{\alpha} = \left(1 - \frac{t}{\beta}\right)^{-\alpha}, t < \beta.$$

Note:

1. The exponential distribution is a special case of gamma distribution with $\frac{1}{\beta}$ parameter when $\alpha = 1$.
2. $\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}$.

1.3.4 Weibull Distribution

The continuous random variable X has a Weibull distribution, with parameters α and $\frac{1}{\beta}$, if its pdf is given by

$$f\left(x; \alpha, \frac{1}{\beta}\right) = \alpha\beta x^{\beta-1} e^{-\alpha x^\beta}, \quad x \geq 0$$

where $\alpha > 0$ and $\beta > 0$.

The cumulative distribution function for the Weibull distribution is given by

$$F(x) = 1 - e^{-\alpha x^\beta}.$$

Note: For $\beta = 1$, the Weibull density reduces to the exponential density function.

1.3.5 Chi-Squared Distribution

The random variable X has a chi-squared distribution with $v > 0$ degrees of freedom, denoted as, $X \sim \chi^2(v)$, if its pdf is given by

$$f(x, v) = \frac{1}{2^{\left(\frac{v}{2}\right)}\Gamma\left(\frac{v}{2}\right)} x^{\frac{v}{2}-1} e^{-\frac{x}{2}}, \quad x > 0$$

The mean and the variance are

$$E(X) = v \qquad \text{and} \qquad V(x) = 2v.$$

The mgf of this distribution is $M(t) = (1-2t)^{-\frac{v}{2}}, t > \frac{1}{2}$.

Note: It is a special case of gamma distribution in which $\alpha = \frac{v}{2}$ and $\beta = \frac{1}{2}$.

Example 1.7:

Let X be a $\chi^2(10)$. Find:

1. Find $P(X > 20.5)$.
2. a, if $P(X > a) = 0.05$.

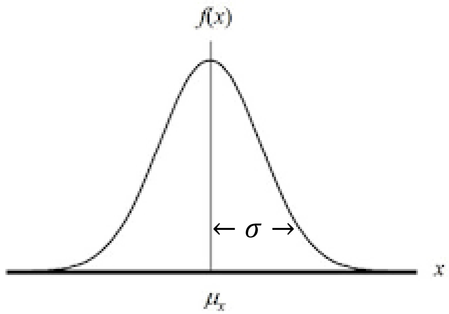
Solution:

By χ^2 Table (Table I) and $v = 10$, we get

1. $P(X > 20.5) = 0.025$
2. $P(X > a) = 0.05$, thus $a = 18.31$.

1.3.6 Normal Distribution

The most important continuous probability distribution in the entire field of statistics is the **normal distribution**. Its graph, called the **normal curve**, is the bell-shaped curve of following figure, which approximately describes many phenomena that occur in nature, industry, and research.



Definition:

The density of the normal random variable X, with mean μ and variance σ^2, $X \sim N(\mu, \sigma^2)$, is

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

where $-\infty < \mu < \infty$ and $\sigma > 0$.

The properties of the normal curves:

1. The mode = median = mean = μ.
2. The curve is symmetric about the mean μ.
3. The normal curve depends on the parameters μ and σ, its mean and standard deviation, respectively.
4. The mean μ and the variance σ^2 determine the location and the shape of the normal curve, respectively.

5. The total area under the curve and above the horizontal axis is equal to 1.
6. The mgf is given by $M(t) = e^{\mu t+\frac{1}{2}\sigma^2 t^2}$.

1.3.7 Standard Normal Distribution

The distribution of a normal random variable with mean 0 and variance 1 is called a **standard normal distribution** and defined as

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

The properties of the standard normal curves:

1. The mode = median = mean = 0.
2. The curve is symmetric about the mean 0.
3. The total area under the curve and above the horizontal axis is equal to 1.
4. The mgf is given by $M(t) = e^{\frac{1}{2}t^2}$.

Application: we are able to transform all the observations of any normal random variable X into a new set of observations of a normal random variable Z with mean 0 and variance 1. This can be done by mean of the transformation i.e.

$$\text{If } X \sim N(\mu, \sigma^2), \text{ then } Z = \frac{X-\mu}{\sigma} \sim N(0,1).$$

Example 1.8:

Given a standard normal distribution, find the area under the curve that lies

1. to the left of $z = 1.84$.
2. to the right of $z = 1.84$.

Solution: From Table II,

1. the area to the left of $z = 1.84$ is equal to,
$$P(Z < 1.84) = 0.9671.$$
2. the area to the right of $z = 1.84$ is equal to,
$$P(Z > 1.84) = 1 - P(Z < 1.84) = 1 - 0.9671 = 0.0329.$$

Normal Approximation to the Binomial:

Theorem 1.1:

If X is a binomial random variable with mean $\mu = np$ and variance $\sigma^2 = npq$, then the limiting form of the distribution of

$$Z = \frac{X - np}{\sqrt{npq}} \sim N(0,1)$$

as $n \to \infty$.

1.3.8 *T*-Distribution

A continuous random variable T is said to have a t-distribution with parameter $\nu > 0$ if its pdf defined as

$$f(t;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}; \quad -\infty < t < \infty.$$

The properties of the standard normal curves:

1. The mode = median = mean = 0.
2. The curve is symmetric about the mean 0.
3. Compared to the standard normal distribution, the t-distribution is less peaked in the center and has higher tails.
4. It depends on the degrees of freedom ν.
5. T-distribution approaches the standard normal distribution as $\nu \to \infty$.
6. The total area under the curve and above the horizontal axis is equal to 1.

Example 1.9: Find:

1. $P(T < 2.145)$ when $\nu = 14$.
2. $t_{0.995}$ when $\nu = 7$.

Solution: From Table III,

1. $P(T < 2.145) = 0.975$ when $\nu = 14$.
2. $t_{0.995} = 3.499$ when $\nu = 7$.

1.3.9 *F*-Distribution

If a random variable *X* has a *F*-distribution with parameters *r* and *v*, we write $X \sim F(r, v)$. Then the probability density function for *X* is given by

$$f(x; r, v) = \frac{1}{B\left(\frac{r}{2}, \frac{v}{2}\right)} \left(\frac{r}{v}\right)^{\frac{r}{2}} x^{\frac{r}{2}-1} \left(1 + \frac{r}{v} x\right)^{-\left(\frac{r+v}{2}\right)}$$

For real $x \geq 0$. Here is $B(a, b) = \int_0^1 y^{a-1}(1-y)^{b-1} dy$ is the beta function and $r, v > 0$.

Theorem 1.2:

If $F_\alpha(r, v)$ has *F*-distribution with *r* and *v* degrees of freedom, then

$$F_{1-\alpha}(v, r) = \frac{1}{F_\alpha(r, v)}$$

has *F*-distribution with *v* and *r* degrees of freedom.

1.4 Transformation of Variables

In standard statistical methods, the result of statistical hypotheses testing, estimation, or even statistical graphics does not involve a single random variable but, rather, **functions of one or more random variables**. As a result, statistical inference requires the distribution of these functions. In this section, we represent methods to find the distribution of these functions.

1.4.1 Discrete Random Variable

1.4.1.1 One-to-One Transformation:

Theorem 1.3:

Suppose that X is a discrete random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a one-to-one transformation between the values of X and Y so that the equation $y = u(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is

$$g(y) = f[w(y)].$$

Example 1.11:

Let X be a discrete random variable with pmf as

$$f(x) = \frac{x}{4}\,, \quad x = 0, 1, 3.$$

Find the pmf of the random variable $Y = X^2$.

Solution:

Since the value of X are all positive, the transformation defines a one-to-one correspondence between the x and y values.

Hence,

Since $x = 0, 1, 3 \Longrightarrow y = 0, 1, 9$ and $y = x^2 \Longrightarrow x = \sqrt{y}$.

Then, the pmf of Y is given by

$$g(y) = f\left(\sqrt{y}\right) = \frac{\sqrt{y}}{4}, \quad y = 0, 1, 9.$$

Similarly, for a two-dimension transformation.

Theorem 1.4:

Suppose that X_1 and X_2 are discrete random variables with joint probability distribution $f(x_1, x_2)$. Let $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations

$$y_1 = u_1(x_1, x_2) \text{ and } y_2 = u_2(x_1, x_2)$$

may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2\,(y_1, y_2)]$$

1.4.2 Continuous Random Variable

This section introduced three methods of transformation to find the distribution of continuous random variable.

1.4.2.1 One-to-One Transformation

Theorem 1.5:

Suppose that X is a continuous random variable with probability distribution $f(x)$. Let $Y = u(X)$ define a one-to-one correspondence between the values of X and Y so that the equation $y = u(x)$ can be uniquely solved for x in terms of y, say $x = w(y)$. Then the probability distribution of Y is

$$g(y) = f[w(y)].|J|$$

where $|J| = |w'(y)| = \left|\frac{\partial x}{\partial y}\right|$ and is called the **Jacobian** of the transformation.

Example 1.12:

Let X be a continuous random variable with probability distribution

$$f(x) = \begin{cases} \frac{x}{12}, & 1 < x < 5, \\ 0, & \text{elsewhere.} \end{cases}$$

Find the probability distribution of the random variable $Y = 2X - 3$.

Solution:

The inverter solution of $y = 2x - 3$ yields $x = (y + 3)/2$, from which we obtain $J = w'(y) \ = \frac{dx}{dy} = \frac{1}{2}$.

Therefore,

$$1 < x < 5 \implies 1 < \frac{y+3}{2} < 5 \Rightarrow 2 < y+3 < 10 \Rightarrow -1 < y < 7$$

Using Theorem 1.5, we find the density function of Y to be

$$g(y) = \begin{cases} \frac{(y+3)/2}{12}\left(\frac{1}{2}\right) = \frac{y+3}{48}, & -1 < y < 7 \\ 0, & \text{elswhere} \end{cases}.$$

Theorem 1.6:

Suppose that X_1 and X_2 are continuous random variable with joint probability distribution $f(x_1, x_2)$. Let $Y_1 = u_1(X_1, X_2)$ and $Y_2 = u_2(X_1, X_2)$ define a one-to-one transformation between the points (x_1, x_2) and (y_1, y_2) so that the equations $y_1 = u_1(x_1, x_2)$ and $y_2 = u_2(x_1, x_2)$ may be uniquely solved for x_1 and x_2 in terms of y_1 and y_2, say $x_1 = w_1(y_1, y_2)$ and $x_2 = w_2(y_1, y_2)$. Then the joint probability distribution of Y_1 and Y_2 is

$$g(y_1, y_2) = f[w_1(y_1, y_2), w_2(y_1, y_2)].\,|J|$$

where the Jacobian is 2×2 determinant as

$$|J| = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}.$$

1.4.2.2 Distribution Function Method (cdf Method):

The general method works as follows:

If X be an independent random variable with pdf $f_X(x)$ and $Y = u(X)$ be a function of X. Then, find

1. $F_X(x)$, cdf of X.
2. The region of Y.
3. $F_Y(y) = P(Y \leq y) = P(u(X) \leq y) = P(X \leq w(Y)) = F_X(w(Y))$.
4. The density function $f_Y(y)$ by differentiating $F_Y(y)$.

Example 1.13:

Suppose the random variable X has a pdf

$$f_X(x) = 3x^2, \quad 0 < x < 1.$$

Find the pdf of $Y = 2X + 3$.

Solution:

From Example 1.3, we get $F_X(x) = x^3$.

Since $0 < x < 1 \Longrightarrow 0 < 2x < 2 \Rightarrow 3 < y < 5$.

$$F_Y(y) = P(Y \leq y) = P(2X + 3 \leq y) = P(2X \leq y - 3) = P\left(X \leq \frac{y-3}{2}\right) = F_X\left(\frac{y-3}{2}\right) = \left(\frac{y-3}{2}\right)^3.$$

Then, the pdf of Y is $f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{3}{8}(y-3)^2,\ 3 < y < 5$.

1.4.2.3 Moment-Generating Method:

Theorem 1.7: (uniqueness Theorem)

Let X and Y be two random variables with moment-generating functions $M_X(t)$ and $M_Y(t)$, respectively, if $M_X(t) = M_Y(t)$ for all values of t, then X and Y have the same probability distribution.

Theorem 1.8:

If $X_1, X_2, \dots\dots, X_n$ are independent random variable with moment-generating functions $M_{X_1}(t), M_{X_2}(t), \dots\dots, M_{X_n}(t)$, respectively, and $Y = X_1 + X_2 + \cdots. + X_n$, then

$$M_Y(t) = M_{X_1}(t). M_{X_2}(t) \dots M_{X_n}(t).$$

Moreover, if $M_{X_1}(t), M_{X_2}(t), \dots\dots, M_{X_n}(t)$are equals. Then, $M_Y(t) = \left(M_{X_1}(t)\right)^n$.

Example 1.14:

If $X_1, X_2, \dots\dots, X_n$ are independent, each with an exponential distribution with parameter $\frac{1}{\theta}$. Show that $Y = \sum_{i=1}^{n} X_i$ has a gamma distribution with parameters n and $\frac{1}{\theta}$.

Solution:

Since that the mgf of $expnential\left(\frac{1}{\theta}\right)$ is $M_X(t) = \frac{\theta}{\theta - t}$. Thus, the mgf of Y is given by

$$M_Y(t) = M_{\sum_{i=1}^{n} X_i}(t) = M_{X_1+X_2+\cdots.+X_n}(t) = M_{X_1}(t). M_{X_2}(t) \dots M_{X_n}(t) = \left(\frac{\theta}{\theta - t}\right)^n.$$

which is the mgf of $Gamma\left(n, \frac{1}{\theta}\right)$.

Chapter 2: Sampling Distribution

In a typical statistical problem, we have a random variable X of interest but its probability distribution $f(x)$ is not known. This problem can be classified in one of two ways:

1. $f(x)$ is completely unknown (Sampling Distribution).
2. The form of $f(x)$ is known but the parameter θ is unknown (Statistical Inference).

In this chapter, we will discuss the first problem and introduce some solution methods. First, let us begin with important definitions.

Random sample:

Let $X_1, X_2, \dots, X_n$ be a *n* independent random variables, each of which has the same probability distribution $f(x)$. Define $X_1, X_2, \dots, X_n$ to be a **random sample** of size *n* from the population $f(x)$ and write its joint probability distribution as

$$f(x_1, x_2, \dots, x_n) = f_1(x_1) f_2(x_2) \dots f_n(x_n).$$

Statistic:

Any function of the random sample and <u>does not depend</u> upon any unknown parameter is called a **statistic**.

Sampling Distribution:

The probability distribution of a statistic is called a **sampling distribution**.

In this chapter, we studied several of the important sampling distributions of frequently used statistic. Applications of these sampling distributions to problems of statistical inference are considered throughout most of the remaining chapters.

In Chapter 1 we defined the two parameters μ and σ^2, which measure the center of location and the variability of a probability distribution, respectively. Here, we shall define some important statistics that describe corresponding measures of a random sample. The most common statistics are the sample mean and variance.

Mean and Variance:

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size *n* from a given distribution. The statistic

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

is called the **mean** of the random sample, and the statistic

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

is called the **variance** of the random sample.

Now, we should view the sampling distribution of $\bar{X}$ and S^2 as the mechanisms from which we will be able to make inference on the unknown parameters μ and σ^2.

2.1 Sampling Distribution of $\bar{X}$

Suppose that we have a population with mean μ and variance σ^2 and let $X_1, X_2, \ldots, X_n$ be a random sample of size *n* from this population. Let the mean of the random sample be $\bar{X}$. Now, consider the following theorems of different cases of sampling distribution of $\bar{X}$.

Theorem 2.1:

Let $X_1, \dots, X_n$ be independent random variables such that, for $i = 1, \dots, n$, X_i has a $N(\mu_i, \sigma_i^2)$ distribution. Let $Y = \sum_{i=1}^{n} a_i X_i$, where $a_1, \dots, a_n$ are constants. Then, the distribution of Y is $N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$.

Proof:

Using independent and the mgf of normal distribution, for $t \in R$, the mgf of Y is,

$$M_Y(t) = E(e^{tY}) = E\left[e^{t\sum_{i=1}^{n} a_i X_i}\right]$$

$$= \prod_{i=1}^{n} E[e^{t a_i X_i}] = \prod_{i=1}^{n} e^{a_i \mu_i t + \frac{1}{2} a_i^2 \sigma_i^2 t^2}$$

$$= e^{\sum_{i=1}^{n} a_i \mu_i t + \frac{1}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2 t^2}$$

which is the mgf of a $N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$ distribution.

Example 2.1:

Let $X_1 \sim N(3,2)$ and $X_2 \sim N(2,1)$. Find the distribution of $Y = 5X_1 - 2X_2$.

Solution:

$$E(Y) = 5E(X_1) - 2E(X_2) = 15 - 4 = 11$$

$$Var(Y) = 25E(X_1) + 4E(X_2) = 50 + 4 = 54$$

Then, the distribution of Y is obtained as $Y \sim N(11, 54)$.

Theorem 2.2:

If $X_1, X_2, \dots, X_n$ is a random sample from any distribution with mean μ and variance σ^2; then

$$\mu_{\bar{X}} = \mu \text{ and variance } \sigma_{\bar{x}}^2 = \frac{\sigma^2}{n}.$$

Proof:

Since $X_1, X_2, \dots, X_n$ is a random sample, then

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

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

Theorem 2.3:

Suppose that $X_1, X_2, \dots, X_n$ be a random sample of n observations are taken from a normal population with mean μ and variance σ^2. Each observation $X_i, i = 1, 2, \dots, n$, has the same normal distribution. Hence, we conclude that

1. $\bar{X}$ has a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, $\left[\text{i.e. } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right]$.
2. $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

Proof:

Since, we know $X_1, X_2, \dots, X_n$ are independent random variables and have the same normal distribution, then they have the same mgf which is

$$M_{X_i}(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2}, i = 1, 2, \dots, n.$$

Now, by using the mgf transformation method (Theorem 1.8), we get

$$M_{\bar{X}}(t) = E\left(e^{\bar{X}t}\right) = E\left(e^{\frac{1}{n}\sum_{i=1}^{n} X_i\, t}\right) = E\left(e^{\frac{1}{n}(X_1+X_2+\cdots+X_n)\, t}\right)$$

$$= E\left(e^{X_1 \frac{t}{n} + X_2 \frac{t}{n} + \cdots + X_n \frac{t}{n}}\right) = \left(M_{X_1}\left(\frac{t}{n}\right)\right)^n, \text{ for any random variable } X_1$$

$$= \left(e^{\mu \frac{t}{n} + \frac{1}{2}\sigma^2 \frac{t^2}{n^2}}\right)^n = e^{\mu t + \frac{1}{2}\frac{\sigma^2}{n} t^2}.$$

which is the mgf of the normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Theorem 2.4: Central Limit Theorem:

If $X_1, X_2, \dots, X_n$ is a random sample of size n from any distribution with mean μ and variance σ^2; if $\bar{X}$ is the mean of the random sample, then as $n \to \infty$,

1. $\bar{X}$ has approximately a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$, $\left[\text{i.e. } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)\right]$.
2. $Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$.

Example 2.2:

An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with mean equal to 800 hours and a standard deviation of 40 hours, find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.

Solution:

The sampling distribution of $\bar{X}$ will be approximately normal, with $\mu_{\bar{X}} = 800$ and $\sigma_{\bar{x}} = \frac{40}{\sqrt{16}} = 10$. Then,

$$P(\bar{X} < 775) = P\left(Z < \frac{775-800}{10}\right) = P(Z < -2.5) = 0.0062.$$

Theorem 2.5:

Let $X_1, X_2, \dots, X_n$ is a random sample of size n from a normal distribution with mean μ and <u>unknown</u> variance σ^2, then

1. $\bar{X}$ has a t-distribution with mean μ, variance $\frac{S^2}{n}$ and $(n-1)$ degrees of freedom.
2. $T = \frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$.

Example 2.3:

A sample of 16 ten-year-old girls had a standard deviation of 12 pounds. Assume the population is normal distribution with mean weight 70 pounds. Find $P(\bar{X} > 74)$.

Solution:

We have, $\mu = 70$, $S = 12$ and $n = 16$. Then, $\bar{X}$ has a t-distribution with $n-1 = 15$ degree of freedom. Thus,

$$P(\bar{X} > 74) = 1 - P\left(T < \frac{74-70}{12/\sqrt{16}}\right) = 1 - P(T < 1.333) = 1 - 0.9 = 0.1$$

2.2 Sampling Distributions from the Normal and Chi-Squared Distributions

In this section we introduce some sampling distributions of some important and useful random variables.

Theorem 2.6:

Let $Z \sim N(0,1)$. Then, $U = Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2$ follows the chi-squared distribution with 1 degree of freedom i.e. $Z^2 \sim \chi_1^2$.

Proof:

We know that the pdf of Z is $f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$. Now, to find the distribution of U, use the cdf transformation method as following:

$$F_U(u) = P(U \le u) = P(Z^2 \le u) = P(-\sqrt{u} \le Z \le \sqrt{u}) = F_Z(\sqrt{u}) - F_Z(-\sqrt{u}).$$

Therefore,

$$\begin{aligned} f_U(u) &= f_Z(\sqrt{u})\frac{dz}{du} - f_Z(-\sqrt{u})\frac{dz}{du} \\ &= \frac{1}{2}u^{-\frac{1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u} + \frac{1}{2}u^{-\frac{1}{2}}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}u} = \frac{1}{2^{\frac{1}{2}}\Gamma\left(\frac{1}{2}\right)}u^{-\frac{1}{2}}e^{-\frac{u}{2}}. \end{aligned}$$

which is the pdf of chi-squared distribution with 1 degree of freedom.

Corollary 2.1:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. If the mean of the random sample is $\bar{X}$, where $\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$ and $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$, then

$$\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2.$$

Proof:

Left as an exercise.

Theorem 2.7:

Let $Z_1, Z_2, \dots\dots, Z_n$ be independent random variables with $Z_i = \frac{X_i - \mu_i}{\sigma_i} \sim N(0,1)$, where $X_i \sim N(\mu_i, \sigma_i)$ for each $i = 1, 2, \dots, n$. If $Y = \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left(\frac{X_i - \mu_i}{\sigma_i}\right)^2$ then Y follows the chi-squared distribution with n degrees of freedom. We write $Y = \sum_{i=1}^{n} z_i^2 \sim \chi_n^2$.

Proof:

Since $Z_1, Z_2, \dots, Z_n$ are independent, then

$$\begin{aligned} M_Y(t) = M_{\sum_{i=1}^{n} z_i^2}(t) &= E\left(e^{(z_1^2+z_2^2+\cdots+z_n^2)t}\right) \\ &= E\left(e^{z_1^2 t}\right). E\left(e^{z_2^2 t}\right) \dots E\left(e^{z_n^2 t}\right) \\ &= M_{z_1^2}(t)\, M_{z_2^2}(t) \dots M_{z_n^2}(t) \end{aligned}$$

From Theorem 2.6, each Z_i^2 follows χ_1^2 and therefore it has mgf equal to $(1-2t)^{-\frac{1}{2}}$. Conclusion:

$$M_Y(t) = \left(M_{z_1^2}(t)\right)^n = (1-2t)^{-\frac{n}{2}}, \text{ for } t > \frac{1}{2}$$

This is the mgf of chi-squared distribution with n degrees of freedom.

Corollary 2.2:

Let $X_1, X_2, \dots, X_n$ is a random sample from $N(\mu, \sigma^2)$, then $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$.

Theorem 2.8:

If $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ is the sample variance of a random sample from a normal distribution with mean μ and variance σ^2, then

$$U = \frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Proof:

Since $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$; then we can redefine U as

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

Now, let

$$\begin{aligned}
\sum_{i=1}^{n}(X_i - \bar{X})^2 &= \sum_{i=1}^{n}[(X_i - \mu) - (\bar{X} - \mu)]^2 \\
&= \sum_{i=1}^{n}[(X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2] \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(n\bar{X} - n\mu)(\bar{X} - \mu) + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2n(\bar{X} - \mu)^2 + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2.
\end{aligned}$$

Then,

$$U = \frac{1}{\sigma^2}\left[\sum_{i=1}^{n}(X_i - \mu)^2 - n(\bar{X} - \mu)^2\right] = \sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2.$$

Use the mgf transformation method to find the distribution of U as follows

$$M_U(t) = E(e^{U\,t}) = E\left(e^{\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 - \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2\right]t}\right)$$

Since, $\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2$ and $\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$ are independent random variables (**Prove it**), we can get

$$M_U(t) = E\left(e^{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 t}\right)E\left(e^{-\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 t}\right) = \frac{E\left(e^{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 t}\right)}{E\left(e^{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 t}\right)} = \frac{M_{\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}(t)}{M_{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}(t)}$$

From Corollary 2.1 and Corollary 2.2, we found that

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2 \quad \text{and} \quad \left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2 \sim \chi_1^2$$

i.e.,

$$M_{\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2}(t) = (1-2t)^{-\frac{n}{2}} \text{ and } M_{\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2}(t) = (1-2t)^{-\frac{1}{2}}$$

Then,

$$M_U(t) = \frac{(1-2t)^{-\frac{n}{2}}}{(1-2t)^{-\frac{1}{2}}} = (1-2t)^{-\frac{(n-1)}{2}}$$

which is the mgf of chi-squared distribution with $n-1$ degrees of freedom. Thus,

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

Theorem 2.9:

Let $X \sim \chi_n^2$, $Y \sim \chi_m^2$. If X, Y are independent then $X + Y \sim \chi_{n+m}^2$.

Proof:

Left as an exercise.

Theorem 2.10:

Let Z denote a random variable that is $Z \sim N(0,1)$; let U denote a random variable that is $U \sim \chi_k^2$ and let Z and U are independent. Then,

$$T = \frac{Z}{\sqrt{U/k}} \sim t_k$$

Proof:

Since are Z and U independent, the joint density of Z and U is given by

$$f_{Z,U}(z,u) = f_Z(z).f_U(u)$$

$$= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)} u^{\frac{k}{2}-1} e^{-\frac{u}{2}}$$

$$= \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)\sqrt{2\pi}} u^{\frac{k}{2}-1} e^{-\frac{1}{2}z^2-\frac{u}{2}}, \quad u > 0, -\infty < z < \infty$$

The one-to-one transformation will be used to obtain the pdf of T. Define the random variables

$$T = \frac{Z}{\sqrt{U/k}} \text{ and } Y = U$$

Then, we can write

$$z = \frac{t\sqrt{y}}{\sqrt{k}} \text{ and } u = y$$

Therefore, the Jacobian is

$$|J| = \begin{vmatrix} \frac{\partial z}{\partial t} & \frac{\partial z}{\partial y} \\ \frac{\partial u}{\partial t} & \frac{\partial u}{\partial y} \end{vmatrix} = \begin{vmatrix} \frac{\sqrt{y}}{\sqrt{k}} & \frac{t}{2\sqrt{ky}} \\ 0 & 1 \end{vmatrix} = \frac{\sqrt{y}}{\sqrt{k}}.$$

Thus, the joint pdf of T and Y is given by

$$f_{T,Y}(t,y) = f_{Z,U}\left(\frac{t\sqrt{y}}{\sqrt{k}}, y\right).|J| = \frac{1}{2^{\left(\frac{k}{2}\right)}\Gamma\left(\frac{k}{2}\right)\sqrt{2\pi}}\, y^{\frac{k}{2}-1} e^{-\frac{yt^2}{2k}-\frac{y}{2}} \frac{\sqrt{y}}{\sqrt{k}}, \quad y > 0, -\infty < t < \infty$$

The marginal pdf of T is then

$$f_T(t) = \int_0^{\infty} f_{T,Y}(t,y)\,dy = \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}} \int_0^{\infty} y^{\frac{k+1}{2}-1} e^{-\frac{y}{2}\left(1+\frac{t^2}{k}\right)}\,dy$$

By using gamma function, $\frac{\Gamma(\alpha)}{\beta^{\alpha}} = \int_0^{\infty} x^{\alpha-1} e^{-\beta x}\,dx$, then we get

$$f_T(t) = \frac{1}{2^{\frac{k+1}{2}}\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}} \frac{\Gamma\left(\frac{k+1}{2}\right)}{\left(\frac{1+\frac{t^2}{k}}{2}\right)^{\frac{k+1}{2}}} = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\Gamma\left(\frac{k}{2}\right)\sqrt{\pi k}}\left(1 + \frac{t^2}{k}\right)^{-\frac{k+1}{2}}; \quad -\infty < t < \infty.$$

And this is the pdf of t-distribution with k degrees of freedom.

Theorem 2.11:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a $N(\mu, \sigma^2)$, where σ^2 is unknown. Then,

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Proof:

Since $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$, write

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} = \frac{(\bar{X}-\mu)/\left(\sigma/\sqrt{n}\right)}{\sqrt{\frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{(n-1)\sigma^2}}}$$

From Theorem 2.3 and Theorem 2.8, we obtain

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \text{ and } \frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{\sigma^2} \sim \chi^2_{n-1}$$

Then, from Theorem 2.10, we conclude that

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

Theorem 2.12:

Let U and V are two independent random variables such that $U \sim \chi^2_n$ and $V \sim \chi^2_m$. Then,

$$\frac{U/n}{V/m} \sim F_{n,m}$$

where n and m are the degrees of freedom of F-distribution.

2.3 Sampling Distribution of S^2

The sample variance S^2 is given by

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

From Theorem 2.8, we found that the distribution of S^2 is

$$\frac{(n-1)S^2}{\sigma^2} \sim X_{n-1}^2$$

By using this conclusion, we can calculate the mean and the variance of S^2 as follows

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

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

Corollary 2.3:

The general derivation of the mean and the variance of the sample variance S^2 that does not assume normality are given by

$$E(S^2) = \sigma^2 \text{ and } Var(S^2) = \frac{\mu_4}{n} - \frac{\sigma^4(n-3)}{n(n-1)}$$

where $\mu_4 = E[(X-\mu)^4]$ is the fourth central moment of X.

2.4 Sampling Distribution of Order Statistics

In this section, the concept of order statistic will be defined and some of their properties.

Order Statistic:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a cumulative distribution function $F(x)$. Then, $Y_1 \leq Y_2 \leq \cdots \leq Y_n$, where Y_i are the X_i arranged in order of increasing degrees and are defined to be the order statistics corresponding to the random sample $X_1, X_2, \dots, X_n$.

Theorem 2.13:

Let $X_1, X_2, \dots, X_n$ be a random sample of size n from a continuous cdf $F(x)$ and pdf $f(x)$; let $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ be the order statistics of this random sample. Then, the marginal pdf of any order statistic of order k, say Y_k is given by

$$f_{Y_k}(y_k) = \frac{n!}{(k-1)!(n-k)!}[F(y_k)]^{k-1}[1 - F(y_k)]^{n-k} f(y_k), \text{ for } a < y_k < b.$$

Corollary 2.4:

As a result of Theorem 2.13, the marginal pdf of $Y_1 = min[X_1, X_2, \dots, X_n]$ and the marginal pdf of $Y_n = max[X_1, X_2, \dots, X_n]$ are, respectively, given by

$$f_{Y_1}(y_1) = n[1 - F(y_1)]^{n-1} f(y_1), \text{ for } a < y_1 < b$$

$$f_{Y_n}(y_n) = n[F(y_n)]^{n-1} f(y_n), \text{ for } a < y_n < b.$$

Theorem 2.14:

Let $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ be the order statistics based on the random sample $X_1, X_2, \ldots, X_n$ from a continuous distribution with pdf $f(x)$ and support (a, b). Then, the joint pdf of the order statistics is given by,

$$f(y_1, y_2, \ldots, y_n) = n!\, f(y_1) f(y_2) \ldots f(y_n), \quad \text{for } a < y_1 < y_2 < \cdots < y_n < b.$$

Theorem 2.15:

Let $Y_1 \leq Y_2 \leq \cdots \leq Y_n$ be the order statistics based on the random sample $X_1, X_2, \ldots, X_n$. Then, the joint pdf of any two order statistics, say $Y_r < Y_k$, is expressed in terms of cdf $F(x)$ and pdf $f(x)$ as follows

$$f_{r,k}(y_r, y_k) = \frac{n!}{(r-1)!\,(k-r-1)!\,(n-k)!} [F(y_r)]^{r-1} [F(y_k) - F(y_r)]^{k-r-1}$$
$$[1 - F(y_k)]^{n-k} f(y_r) f(y_k), \; a < y_r < y_k < b$$

Example 2.4:

Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from a distribution having pdf

$$f(x) = 2x, \qquad 0 < x < 1$$

Compute:

1. $P\left(\frac{1}{2} < Y_3\right)$.
2. The joint distribution of Y_1 and Y_3.

Solution:

Here $F(x) = x^2$, provided that $0 < x < 1$, so that

1. $f_{Y_3}(y_3) = \frac{4!}{2!1!}(y_3^2)^2(1 - y_3^2)(2y_3) = 24\left(y_3^5 - y_3^7\right),\ 0 < y_3 < 1$

Thus,

$$P\left(\frac{1}{2} < Y_3\right) = \int_{\frac{1}{2}}^{1} f_{Y_3}(y_3)\, dy_3 = \int_{\frac{1}{2}}^{1} 24\left(y_3^5 - y_3^7\right) dy_3 = \frac{243}{256}.$$

2. $f_{1,3}(y_1, y_3) = \frac{4!}{0!1!\,1!}[y_1^2]^0\ [y_3^2 - y_1^2]^1\ [1 - y_3^2]^1\ \ 2y_1\ \ 2y_3$

$$= 96\ y_1\ y_3\ \ [y_3^2 - y_1^2]\quad [1 - y_3^2]$$

Chapter 3: Point Estimation

In this chapter, we begin 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.

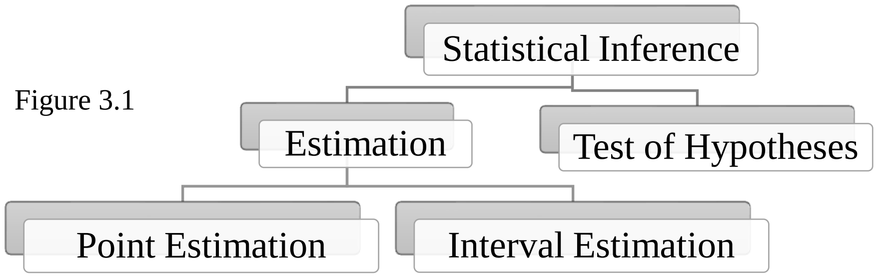


Figure 3.1

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

$$\mu_i' = M_i$$

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

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 $X_1, \dots.., X_n$ 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 $f(x;\ \theta_1, \theta_2, \dots\dots, \theta_r), (\theta_1, \dots, \theta_r) \in \Omega$, the maximum likelihood estimators $\hat{\theta}_1, \hat{\theta}_2, \dots\dots, \hat{\theta}_r$ are that which maximizes the likelihood function $L(\theta_1, \theta_2, \dots\dots, \theta_r; x)$.

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

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

Supposed that a random sample $X_1, X_2, ..., X_n$ 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

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

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: ($n \to \infty$)**
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 $X_1, \dots, X_n$ 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

$$MSE(T) = E[(T-\theta)^2] = Var(T) + \big(\theta - E(T)\big)^2$$

The term $\big(\theta - E(T)\big)$ 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:

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

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

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

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

Theorem 3.1:

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

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

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.

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

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 $X_1, X_2, \dots, X_n$ 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 $X_1, X_2, \dots., X_n$ 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

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

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

Theorem 3.4:

Let $X_1, X_2, \dots., X_n$ 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:

We know that the mean and the variance of X are

$$E(X) = \beta \text{ and } Var(X) = \beta^2$$

Then,

$$E(\bar{X}) = E(X) = \beta.$$

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

$$Var(\bar{X}) = \frac{Var(X)}{n} = \frac{\beta^2}{n}.$$

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 $\bar{X}$ and 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

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

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

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

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:

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

Thus,

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

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

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

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

$$Var(Y_n) = E(Y_n^2) - \left(E(Y_n)\right)^2 = \frac{n\,\theta^2}{(n+2)} - \frac{n^2\,\theta^2}{(n+1)^2} = \frac{n\,\theta^2}{(n+2)(n+1)^2}$$

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 $X_1, X_2, \ldots, X_n$ 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

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

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

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

We observe that we may maximum by differentiation $\ln L(\mu,\sigma^2;x_1,\ldots\ldots,x_n)$ 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 must 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

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

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

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. $\lim\limits_{n\to\infty} Var(\bar{X}) = \lim\limits_{n\to\infty} \frac{\sigma^2}{n} = 0.$

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

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

d) Sufficiency:

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

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

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

$$= \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\left[nS_1^2+n(\bar{X}-\mu)^2\right]}$$

Let $T_1 = \bar{X}, T_2 = S_1^2$. 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 $K_1(T_1, T_2, \mu, \sigma^2) = \left(\sqrt{2\pi}\sigma\right)^{-n} e^{-\frac{1}{2\sigma^2}\left[nT_2+n(T_1-\mu)^2\right]}$ 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

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

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 3.7.

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 3.7, 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

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

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

$$\Rightarrow g(0) = g(1) = \cdots = g(n) = 0 \Rightarrow g(T) = 0, \text{ for all } T.$$

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 $X_1, X_2, \dots, X_n$ 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:

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

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

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

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

2. **Variance (Fisher Information)**

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

Proof:

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

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

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

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

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

The first term in the right side can be written as

$$E\left[\frac{L''}{L}\right] = \int_{x_1} \dots \int_{x_n} \frac{L''}{L} L \, dx_n \dots dx_1$$
$$= \frac{\partial^2}{\partial\theta^2}\int_{x_1} \dots \int_{x_n} L(x,\theta)\, dx_n \dots dx_1 = \frac{\partial^2}{\partial\theta^2}(1) = 0$$

The second term is obtained as

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

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:

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

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,

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

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

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

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

(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

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

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

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

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

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

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

Now consider the covariance of T and score function as following

$$Cov\left[T, \frac{\partial}{\partial\theta}\log L(x;\theta)\right] = E\left[T\frac{\partial}{\partial\theta}\log L(x;\theta)\right] - E[T]E\left[\frac{\partial}{\partial\theta}\log L(x;\theta)\right] = \acute{\tau}(\theta)$$

Now by the Cauchy-Schwarz inequality, we get

$$\left|Cov\left[T, \frac{\partial}{\partial\theta}\log L(x;\theta)\right]\right|^2 \le Var[T]Var\left[\frac{\partial}{\partial\theta}\log L(x;\theta)\right]$$

Then,

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

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 $X_1, X_2, \ldots, X_n$ 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

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

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

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

Remark: $\bar{X}$ 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 $X_1, X_2, \ldots, X_n$ 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

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

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

$$\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2 \sim \chi_n^2$$

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

$$E\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]=n \Rightarrow E\left[\frac{nS_2^2}{\sigma^2}\right]=n \Rightarrow \frac{n}{\sigma^2}E(S_2^2)=n \Longrightarrow E(S_2^2)=\sigma^2$$

$$Var\left[\sum_{i=1}^{n}\left(\frac{X_i-\mu}{\sigma}\right)^2\right]=2n \Longrightarrow Var\left[\frac{nS_2^2}{\sigma^2}\right]=2n \Longrightarrow \frac{n^2}{\sigma^4}Var[S_2^2]=2n \Longrightarrow Var(S_2^2)=\frac{2\sigma^4}{n}$$

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.

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 p. 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

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

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

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

$$= \frac{p \binom{n-1}{S-1} p^{S-1} q^{n-S}}{\binom{n}{S} p^S q^{n-S}} = \frac{(n-1)!}{(S-1)!(n-S)!} \frac{S!(n-S)!}{n!} = \frac{S}{n} = \bar{X}$$

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 $X_1, \dots, X_n$ be a random sample from $f(x, \theta)$, θ may be a vector of parameters $(\theta_1, \dots, \theta_k)$. If $S = s(S_1, \dots, S_k)$ 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, $Var[T^*] \le Var[T]$ for all $\theta \in \emptyset$; so $T^*$is an UMVUE.

Example 3.17:

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

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

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

$$E(T_1^*) = \beta \Rightarrow E(cS) = \beta \ \Rightarrow c\, E\left(\sum_{i=1}^{n} X_i\right) = \beta \Rightarrow c\, n\beta = \beta \Rightarrow c = \frac{1}{n}$$

   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:** $\lim\limits_{n\to\infty} E(\hat{\theta}) = \theta$
4. **Consistency:** $\lim\limits_{n\to\infty} P(|\hat{\theta} - \theta| \geq \varepsilon) = 0\,,\ \forall\theta$
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

$$\sqrt{n}(\hat{\tau}(\theta) - \tau(\theta)) \xrightarrow{d} N\left(0, \frac{(\tau'(\theta))^2}{I(\theta)}\right) \text{ or } \hat{\tau}(\theta) \xrightarrow{d} N\left(\tau(\theta), \frac{(\tau'(\theta))^2}{nI(\theta)}\right).$$

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 $\theta \longrightarrow \theta$ 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 $\theta \longrightarrow \theta$ 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

$$t(c\,x_1 + d, \dots, c\,x_n + d) = c\,t(x_1, \dots, x_n) + d \text{ for all values } c > 0 \text{ and } d.$$

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

$$t(c\,x_1 + d, \dots, c\,x_n + d) = t(x_1, \dots, x_n) \text{ for all values } c > 0 \text{ and } d.$$

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.

Chapter 4: Interval Estimation

Chapter 3 dealt with the point estimation of a parameter or made the inference of estimating the true value of the parameter to be a point. In this chapter, we might make the inference of estimating that true value of the parameter is contained in some interval that is called interval estimation.

Confidence Interval:

Let $X_1, X_2, \dots, X_n$ be a random sample from $f(x, \theta)$. Let $T_1 = t_1(X_1, \dots, X_n)$ and $T_2 = t_2(X_1, \dots, X_n)$ be two statistics satisfying $T_1 < T_2$ for which $P(T_1 < \tau(\theta) < T_2) = 1 - \alpha$, where α does not depend on θ, then the random interval (T_1, T_2) is called 100 $(1 - \alpha)$% **confidence interval** for $\tau(\theta)$, α is called the confidence coefficient and T_1 and T_2 are called the lower and upper confidence limits, respectively, for $\tau(\theta)$.

4.1 Confidence Intervals from Normal Distribution

In this section, we derive confidence intervals for the mean μ and the variance σ^2 when the random sample $X_1, X_2, \dots, X_n$ has normal distribution.

4.1.1 Confidence Interval for the Mean

There are two cases to consider depending on whether or not σ^2 is known.

First Case (σ^2 is known):

If the sample is selected from a normal population or, if n is large enough, (Theorem 2.3 and Theorem 2.4) the sampling distribution of the sample mean $\bar{X}$ when σ^2 is known is given by

$$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

Then, we establish a 100 $(1-\alpha)$% confidence interval for μ when σ^2 is known as following:

$$P\left(-z_{1-\frac{\alpha}{2}} < \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$P\left(\bar{X} - z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} < \mu < \bar{X} + z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

where $z_{1-\frac{\alpha}{2}}$ is a value from z-table.

Second Case (σ^2 is unknown and n<30):

Now, we turn to the problem of finding a confidence interval for the mean μ of a normal distribution when we are not known the variance σ^2 and the sample size n is small. In Theorem 2.11 we found that

$$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

where S is the sample standard deviation. Then, we can find 100 $(1-\alpha)$% confidence interval for μ when σ^2 is unknown as following:

$$P\left(-t_{(1-\frac{\alpha}{2},n-1)} < \frac{\bar{X}-\mu}{S/\sqrt{n}} < t_{(1-\frac{\alpha}{2},n-1)}\right) = 1-\alpha$$

$$P\left(\bar{X} - t_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{(1-\frac{\alpha}{2},n-1)}\frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

where $t_{(1-\frac{\alpha}{2},n-1)}$ is a value from t-table with $n-1$ degrees of freedom.

Example 4.1:

Let $X_1, X_2, \dots, X_{10}$ be a random sample from $N(\mu, 16)$ and let the sample mean $\bar{X}$ be 3.67. Find 95% confidence interval for the population mean μ.

Solution:

Since population variance is known, $\sigma^2 = 16$, and $\bar{X} = 3.67$, $n = 10$; then 95% confidence interval for the population mean μ is

$$3.67 \pm z_{1-\frac{\alpha}{2}}\frac{4}{\sqrt{10}}$$

where, the value of z-table $z_{1-\frac{\alpha}{2}}$ is found as

$$1 - \alpha = 0.95 \Longrightarrow \alpha = 0.05 \Longrightarrow \frac{\alpha}{2} = 0.025 \Longrightarrow 1 - \frac{\alpha}{2} = 0.975$$

$$\Longrightarrow z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$$

Then,

$$3.67 \pm 1.96\frac{4}{\sqrt{10}} \Longrightarrow 3.67 \pm 2.4792$$

$$\Longrightarrow \mu \in (1.1908, 6.1492)$$

4.1.2 Confidence Interval for the Variance

Let the random variable X be $N(\mu, \sigma^2)$. We shall discuss the problem of finding a confidence interval for σ^2. Our discussion will consist of two parts: the first when μ is a know number, and second when μ is unknown.

First Case (μ is known):

Let $X_1, X_2, \dots, X_n$ denote a random sample of size *n* from distribution that is $N(\mu, \sigma^2)$, where μ is known. From Corollary 2.2, we got that

$$\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} \sim \chi_n^2$$

Let us select a probability, say $1-\alpha$, then 100 $(1-\alpha)$% confidence interval for σ^2 when μ is known is given by

$$P\left(\chi^2_{\left(1-\frac{\alpha}{2},n\right)} < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\sigma^2} < \chi^2_{\left(\frac{\alpha}{2},n\right)}\right) = 1-\alpha$$

$$P\left(\frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\left(\frac{\alpha}{2},n\right)}} < \sigma^2 < \frac{\sum_{i=1}^{n}(X_i-\mu)^2}{\chi^2_{\left(1-\frac{\alpha}{2},n\right)}}\right) = 1-\alpha$$

where $\chi^2_{\left(\frac{\alpha}{2},n\right)}$ and $\chi^2_{\left(1-\frac{\alpha}{2},n\right)}$ are χ^2 values with n degrees of freedom.

Second Case (μ is unknown):

Now, we discuss the case when μ is not known. This case can be handled by making use of the facts from Theorem 2.8 that

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

when the sample variance s^2 is computed. Then, for a fixed positive integer $n \geq 2$, we can find a 100 $(1-\alpha)\%$ confidence interval for σ^2 as

$$P\left(\chi^2_{(1-\frac{\alpha}{2},n-1)} < \frac{(n-1)S^2}{\sigma^2} < \chi^2_{(\frac{\alpha}{2},n-1)}\right) = 1-\alpha$$

$$P\left(\frac{(n-1)S^2}{\chi^2_{(\frac{\alpha}{2},n-1)}} < \sigma^2 < \frac{(n-1)S^2}{\chi^2_{(1-\frac{\alpha}{2},n-1)}}\right) = 1-\alpha$$

where $\chi^2_{(\frac{\alpha}{2},n-1)}$ and $\chi^2_{(1-\frac{\alpha}{2},n-1)}$ are χ^2 values with $n-1$ degrees of freedom.

Example 4.2:

Let $X_1, X_2, \ldots, X_{25}$ be a random sample from normal distribution when the sample variance is equal to 2.3. Find 90% confidence interval for the population variance σ^2.

Solution:

We want to construct a confidence interval for σ^2 when the population is normal with unknown mean, thus we should use the following case

$$P\left(\frac{24(2.3)}{\chi^2_{(\frac{\alpha}{2},24)}} < \sigma^2 < \frac{24(2.3)}{\chi^2_{(1-\frac{\alpha}{2},24)}}\right) = 0.9$$

$$P\left(\frac{55.2}{\chi^2_{(\frac{\alpha}{2},24)}} < \sigma^2 < \frac{55.2}{\chi^2_{(1-\frac{\alpha}{2},24)}}\right) = 0.9$$

$$1-\alpha=0.9 \Longrightarrow \alpha=0.1 \Longrightarrow \frac{\alpha}{2}=0.05 \Longrightarrow 1-\frac{\alpha}{2}=0.95$$

$$\Longrightarrow \chi^2_{(0.05,24)}=36.42 \quad \text{and} \quad \chi^2_{(0.95,24)}=13.85$$

$$\Longrightarrow \sigma^2 \in \left(\frac{55.2}{36.42},\frac{55.2}{13.85}\right) \Longrightarrow \sigma^2 \in (1.5157, 3.9856)$$

4.2 Pivotal Quantity Method

Pivotal Quantity:

Let $X_1, X_2, \dots, X_n$ be a random sample from $f(x,\theta)$. Let $Q = q(X_1, \dots, X_n;\ \theta)$ be a function of $X_1, \dots, X_n$ and θ. If Q has a distribution that does not depend on θ, then Q is defined to be a **pivotal quantity**.

Example 4.4:

Let $X_1, X_2, \dots, X_n$ be a random sample from $N(\theta, 9)$. Then,

1. $\bar{X}-\theta \sim N\left(0, \frac{9}{n}\right)$ and $\frac{\bar{X}-\theta}{3/\sqrt{n}} \sim N(0,1)$ are pivotal quantities.
2. $\bar{X}-2\theta \sim N\left(-\theta, \frac{9}{n}\right)$ is not pivotal quantity.

Pivotal Quantity Method:

If $Q = q(X_1, \dots, X_n;\ \theta)$ is apivotal quantity and has a probability distribution, then for any fixed $0 < \alpha < 1$ there will exist q_1 and q_2 such that $q_1 < q_2$ and $P(q_1 < Q < q_2) = 1 - \alpha$

Therefore, we can find 100 $(1-\alpha)$% confidence interval for $\tau(\theta)$ as

$$P(t_1(x_1, \dots, x_n) < \tau(\theta) < t_2(x_1, \dots, x_n)) = 1 - \alpha$$

where t_1 and t_2 are functions of the random sample does not depend on θ.

Remark:

If $X_1, X_2, \dots, X_n$ is a random sample from $f(x, \theta)$, and the corresponding cumulative distribution function $F(x, \theta)$ is continuous in x. Then, a pivotal quantity can be given as

$$Q = -2\sum_{i=1}^{n} \log F(x_i, \theta) \sim \chi^2_{2n}$$

Then, the $(1-\alpha)100$ % confidence interval for $\tau(\theta)$ is given as

$$P\left(\chi^2_{(1-\frac{\alpha}{2}, 2n)} < Q < \chi^2_{(\frac{\alpha}{2}, 2n)}\right) = 1 - \alpha$$

Example 4.5:

If $X_1, \dots, X_n$ be a random sample from the density function

$$f(x) = \theta\, x^{\theta-1}, \quad 0 < x < 1$$

Find a pivotal quantity for θ and use it to construct $100(1-\alpha)$% confidence interval for θ.

Solution:

The CDF of x is given by

$$F(x) = \int_0^x \theta\, x^{\theta-1} dx = x^{\theta}, \quad 0 < x < 1$$

So, the pivotal quantity can be of the form

$$Q = -2\sum_{i=1}^{n} \log x_i^{\theta} = -2\theta \sum_{i=1}^{n} \log x_i$$

where $Q \sim \chi^2_{2n}$, then one can construct $100(1-\alpha)\%$ confidence interval for θ as

$$P\left(\chi^2_{(1-\frac{\alpha}{2},2n)} < Q < \chi^2_{(\frac{\alpha}{2},2n)}\right) = 1-\alpha$$

$$P\left(\chi^2_{(1-\frac{\alpha}{2},2n)} < -2\theta \sum_{i=1}^{n} \log x_i < \chi^2_{(\frac{\alpha}{2},2n)}\right) = 1-\alpha$$

$$P\left(\frac{-\chi^2_{(\frac{\alpha}{2},2n)}}{2\sum_{i=1}^{n} \log x_i} < \theta < \frac{-\chi^2_{(1-\frac{\alpha}{2},2n)}}{2\sum_{i=1}^{n} \log x_i}\right) = 1-\alpha$$

4.3 Large Sample Confidence Interval

From Section 3.3, the MLE $\hat{\theta}$ of θ, has an asymptotic normal distribution when n is large which is given by

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N\left(0, \frac{1}{I(\theta)}\right) \quad \text{or} \quad \hat{\theta} \xrightarrow{d} N\left(\theta, \frac{1}{nI(\theta)}\right)$$

Thus, we can write

$$\frac{\hat{\theta}-\theta}{1/\sqrt{nI(\theta)}} \sim N(0,1)$$

Use the distribution of the MLE $\hat{\theta}$ to construct $100(1-\alpha)\%$ confidence interval for the parameter θ as following:

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\hat{\theta}-\theta}{1/\sqrt{nI(\theta)}} < z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

$$P\left(\hat{\theta} - z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{nI(\theta)}} < \theta < \hat{\theta} + z_{1-\frac{\alpha}{2}}\frac{1}{\sqrt{nI(\theta)}}\right) = 1-\alpha$$

In General:

Since the MLE $\hat{\tau}(\theta)$ of $\tau(\theta)$, has an asymptotic normal distribution when n is large as following:

$$\hat{\tau}(\theta) \xrightarrow{d} N\left(\tau(\theta), \frac{\left(\acute{\tau}(\theta)\right)^2}{nI(\theta)}\right)$$

$$\Rightarrow \frac{\hat{\tau}(\theta)-\tau(\theta)}{\acute{\tau}(\theta)/\sqrt{nI(\theta)}} \sim N(0,1)$$

Then, $100(1-\alpha)\%$ confidence interval for the parameter $\tau(\theta)$ is given by

$$P(-z_{1-\frac{\alpha}{2}} < \frac{\hat{\tau}(\theta)-\tau(\theta)}{\acute{\tau}(\theta)/\sqrt{nI(\theta)}} < z_{1-\frac{\alpha}{2}}) = 1-\alpha$$

$$P\left(\hat{\tau}(\theta) - z_{1-\frac{\alpha}{2}}\frac{\acute{\tau}(\theta)}{\sqrt{nI(\theta)}} < \tau(\theta) < \hat{\tau}(\theta) + z_{1-\frac{\alpha}{2}}\frac{\acute{\tau}(\theta)}{\sqrt{nI(\theta)}}\right) = 1-\alpha$$

Example 4.6:

Let $X \sim Exponential(\beta)$ with large sample size. Construct $100(1-\alpha)\%$ confidence interval for β.

Solution:

The pdf of the exponential distribution with parameter β is defined as

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

We found from Example 3.2 and Example 3.3 that the MLE of β is $\bar{X}$ and it is an unbiased estimator (i.e. $E(\bar{X}\ ) = \beta$). Thus, $\bar{X}$ has an asymptotic normal distribution that is

$$\bar{X} \sim N\left(\beta, \frac{1}{nI(\theta)}\right)$$

Now we need to derive the Fisher information, $I(\theta)$:

$$\log f(x;\theta) = -\log\beta - \frac{x}{\beta}$$

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

$$I(\theta) = Var\left[\frac{\partial}{\partial\theta}\log f(x;\theta)\right] = Var\left[-\frac{1}{\beta}+\frac{X}{\beta^2}\right] = \frac{Var(X)}{\beta^4} = \frac{1}{\beta^2}$$

The asymptotic normal distribution of the MLE is

$$\bar{X} \sim N\left(\beta, \frac{\beta^2}{n}\right) \quad \text{or} \quad \frac{\bar{X}-\beta}{\beta/\sqrt{n}} = \sqrt{n}\left(\frac{\bar{X}}{\beta}-1\right) \sim N(0,1)$$

Thus, $100(1-\alpha)\%$ confidence interval for β is obtained as

$$P\left(-z_{1-\frac{\alpha}{2}} < \sqrt{n}\left(\frac{\bar{X}}{\beta}-1\right) < z_{1-\frac{\alpha}{2}}\right) = 1-\alpha$$

$$\Rightarrow -\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} < \frac{\bar{X}}{\beta}-1 < \frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

$$\Rightarrow 1-\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}} < \frac{\bar{X}}{\beta} < 1+\frac{z_{1-\frac{\alpha}{2}}}{\sqrt{n}}$$

$$\Rightarrow \frac{\bar{X}\sqrt{n}}{\sqrt{n}+z_{1-\frac{\alpha}{2}}} < \beta < \frac{\bar{X}\sqrt{n}}{\sqrt{n}-z_{1-\frac{\alpha}{2}}}$$

Chapter 5: Bayesian Estimation

In the last two Chapters 3 and 4, we assumed the random sample came from some known probability distribution $f(x,\theta)$ and we used the classic method to estimate the unknown parameter θ which was some fixed. In this Chapter, we will estimate θ using the Bayesian method which is define the unknown parameter θ as a random variable and has a distribution depending on previous information called prior distribution.

Prior and Posterior Distributions

Consider a random variable X that has a distribution of probability that depends upon the symbol θ, where θ is an element of a well-defined set Ω. Let us now introduce a random variable Θ that has a distribution of probability over the set Ω. The probability distribution $h(\theta)$ is called the **prior distribution** of Θ. Moreover, we now denote the probability distribution of X by $f(x|\theta)$ since we think of it as a conditional distribution of X, given $\Theta = \theta$. For clarity in this chapter, we will use the following summary of this model:

$$X|\theta \sim f(x|\theta)$$

$$\Theta \sim h(\theta)$$

Thus, we can write the joint conditional distribution of X, given $\Theta = \theta$, as

$$L(x|\theta) = f(x_1|\theta)f(x_2|\theta) \dots f(x_n|\theta)$$

Thus, the joint distribution of X and Θ is

$$g(x,\theta) = L(x|\theta)h(\theta) \tag{5.1}$$

The marginal distribution of X is given by

$$g_1(x) = \begin{cases} \int_\theta \; g(x,\theta)d\theta, & \text{if } \Theta \text{ is a continuous} \\ \sum_\theta \; g(x,\theta)\,, & \text{if } \Theta \text{ is a discrete} \end{cases}$$

In either case the conditional distribution of Θ, given the sample X, is

$$k(\theta|x) = \frac{g(x,\theta)}{g_1(x)} = \frac{L(x|\theta)h(\theta)}{g_1(x)} \tag{5.2}$$

The distribution defined by this conditional distribution is called the **posterior distribution**. The prior distribution reflects the subjective belief of Θ before the sample is drawn while the posterior distribution is the conditional distribution of Θ after the sample is drawn. Further discussion on these distributions follows an illustrative example.

Example 5.1:

Consider the model

$$X_i|\theta \;\sim \text{iid } Poisson\ (\theta)$$

$$\Theta \;\sim\; \Gamma(\alpha,\beta), \alpha \text{ and } \beta \text{ are known}$$

Hence, the random sample is drawn from a Poisson distribution with mean θ and the prior distribution is $\Gamma(\alpha,\beta)$ distribution. Thus, in this case, the joint conditional pdf of X, given $\Theta = \theta$, is

$$L(x|\theta) = \frac{\theta^{x_1}e^{-\theta}}{x_1!} \dots . \frac{\theta^{x_n}e^{-\theta}}{x_n!}, x_i = 0, 1, 2, \dots, i = 1, 2, \dots ., n,$$

and the prior pdf is

$$h(\theta) = \frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}, 0 < \theta < \infty$$

Hence, the joint mixed continuous discrete pdf is given by

$$g(x,\theta) = L(x|\theta)h(\theta) = \left[\frac{\theta^{x_1} e^{-\theta}}{x_1!} \dots \frac{\theta^{x_n} e^{-\theta}}{x_n!}\right]\left[\frac{\theta^{\alpha-1} e^{-\frac{\theta}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}\right]$$

$$= \frac{\theta^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-\left(\frac{n\beta+1}{\beta}\right)\theta}}{\prod_{i=1}^{n} x_i!\, \Gamma(\alpha)\beta^{\alpha}}$$

Provided that $x_i = 0, 1, 2, 3, \dots., i = 1, 2, \dots, n$ and $0 < \theta < \infty$. Then, the marginal distribution of the sample, is

$$g_1(x) = \int_0^{\infty} \frac{\theta^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-\left(\frac{n\beta+1}{\beta}\right)\theta}}{\prod_{i=1}^{n} x_i!\, \Gamma(\alpha)\beta^{\alpha}} d\theta = \frac{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)}{\prod_{i=1}^{n} x_i!\, \Gamma(\alpha)\beta^{\alpha} \left(\frac{n\beta+1}{\beta}\right)^{\sum_{i=1}^{n} x_i + \alpha}}$$

Finally, the posterior pdf of Θ, given $X = x$, is

$$k(\theta|x) = \frac{g(x,\theta)}{g_1(x)} = \frac{\theta^{\sum_{i=1}^{n} x_i + \alpha - 1} e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)\left(\frac{\beta}{n\beta+1}\right)^{\sum_{i=1}^{n} x_i + \alpha}}$$

Provided that $0 < \theta < \infty$. This conditional pdf is one of the gamma type with parameters $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$ and $\beta^* = \frac{\beta}{n\beta+1}$. Notice that the posterior pdf reflects both prior information (α, β) and sample information $(\sum_{i=1}^{n} x_i)$.

$$\theta|X_i \sim \Gamma\left(\sum_{i=1}^{n} x_i + \alpha, \frac{\beta}{n\beta + 1}\right)$$

Remarks:

1. In Example 5.1 notice that it is not really necessary to determine the marginal pdf $g_1(x)$ to find the posterior pdf $k(\theta|x)$. If we divide $L(x|\theta)h(\theta)$ by $g_1(x)$, we must get the product of a factor, which depend upon x but does not depend upon θ, say $c(x)$, That is,

$$k(\theta|x) = c(x)\, \theta^{\sum_{i=1}^{n} x_i + \alpha - 1}\, e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}$$

Provided that $0 < \theta < \infty$, and $x_i = 0, 1, 2, 3, \dots, i = 1, 2, \dots, n$. However, $c(x)$ must be that "constant" needed to make $k(\theta|x)$ a pdf, namely

$$c(x) = \frac{1}{\Gamma\left(\sum_{i=1}^{n} x_i + \alpha\right)\left(\frac{\beta}{n\beta + 1}\right)^{\sum_{i=1}^{n} x_i + \alpha}}$$

Accordingly, we frequently write that $k(\theta|x)$ is proportional to $L(x|\theta)h(\theta)$; that is, the posterior pdf can be written as

$$k(\theta|x) \propto L(x|\theta)h(\theta) \tag{5.3}$$

Note that in the right-hand member of this expression all factors involving constants and x alone (not θ) can be dropped. For illustration, in solving the problem presented in Example 5.1, we simply write

$$k(\theta|x) \propto \theta^{\sum_{i=1}^{n} x_i+\alpha-1} e^{-\frac{\theta}{\left(\frac{\beta}{n\beta+1}\right)}}$$

$0 < \theta < \infty$. Clearly, $k(\theta|x)$ must be gamma pdf with parameter $\alpha^* = \sum_{i=1}^{n} x_i + \alpha$ and $\beta^* = \frac{\beta}{n\beta+1}$.

2. There is another observation that can be made at this point. Suppose that there exists a sufficient statistic $T = t(X)$ for the parameter so that

$$L(x|\theta) = f_T(t \mid \theta).\, k(X),$$

where now $f_T(t \mid \theta)$ is the pdf of T, given $\Theta = \theta$. Then we note that

$$k(\theta|x) \propto f_T(t \mid \theta)\, h(\theta) \tag{5.4}$$

5.1 Bayesian Point Estimation

Suppose we want a point estimator of θ. From the Bayesian viewpoint, this really amounts to selecting a decision function δ, so that $\delta(x)$ is a predicted value of θ (an experimental value of the random variable Θ) when both the computed value x and the conditional pdf $k(\theta|x)$ are known. Now, in general, how would we predict an experimental value of any random variable, say W, if we want our prediction to be "reasonably close" to the value to be observed?. Many statisticians would predict the mean, $E(W)$, of the distribution of W; others would predict a median (perhaps unique) of the distribution of W, and some would have other predictions. However, it seems desirable that the choice of the decision function should depend upon a loss function $\mathcal{L}[\theta, \delta(x)]$. One way in which this dependence upon the loss function can be reflected is to select the decision function δ in such a way that the conditional expectation of the loss is minimum. A **Bayes' estimate** is a decision function δ that minimizes the expectation of the loss function $E\{\mathcal{L}[\Theta, \delta(x)]|X = x\}$ and then

$$\delta(x) = E\{\mathcal{L}[\Theta, \delta(x)]|X = x\} = \begin{cases} \int_\theta \mathcal{L}[\theta, \delta(x)]k(\theta|x)\,d\theta, & \text{if } \Theta \text{ is a continuous} \\ \sum_\theta \mathcal{L}[\theta, \delta(x)]k(\theta|x), & \text{if } \Theta \text{ is a continuous} \end{cases} \tag{5.5}$$

is called **Bayes' estimator** of θ.

Some Possible Loss Functions:

1. **Squared Error Loss Function:**
   The squared error loss function is given by
   $$\mathcal{L}[\theta, \delta(x)] = [\theta - \delta(x)]^2$$
   Then, the Bayes' estimate is the mean of the conditional distribution of Θ, given $X = x$
   $$\delta(x) = E(\Theta|x)$$
2. **Absolute Error Loss Function:**
   The absolute error loss function is given by
   $$\mathcal{L}[\theta, \delta(x)] = |\theta - \delta(x)|$$
   Then, a median of the conditional distribution of Θ, given $X = x$, is the Bayes' solution
   $$\delta(x) = \text{Median of } \Theta$$
   where the median, m, is the solution of
   $$\int_{-\infty}^{m} k(\theta|x)d\theta = \frac{1}{2}$$

It is easy to generalize this to estimate a function of θ, for a specified function $\tau(\theta)$. For the loss function $\mathcal{L}[\theta, \delta(x)]$, a **Bayes estimate** of $\tau(\theta)$ is a decision function δ that minimizes

$$E\{\mathcal{L}[\tau(\Theta), \delta(x)]|X = x\} = \int_{-\infty}^{\infty} \mathcal{L}[\tau(\theta), \delta(x)]\,k(\theta|x)d\theta \tag{5.6}$$

The random variable $\delta(X)$ is called **Bayes' estimator** of $\tau(\theta)$.

Example 5.2:

Consider the model

- $X_i|\theta \sim$ iid $Binomial\ (1,\theta)$

$\Theta \sim Beta(\alpha,\beta)$, α and β are known

That is, the prior pdf is

$$h(\theta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}\ ,0<\theta<1$$

when α and β are assigned positive constants. We seek a decision function δ that is a Bayes' solution. The sufficient statistic is $Y = \sum_1^n X_i$, which has a $Binomial\ (n,\theta)$ distribution. Thus, the conditional pdf of Y given $\Theta = \theta$ is

$$g(y|\theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y} \quad y = 0,1,\dots,n$$

Thus by Equation (5.4), the conditional posterior pdf of Θ, given $Y = y$ at positive probability density, is

$$k(\theta|y) \propto \theta^y(1-\theta)^{n-y}\,\theta^{\alpha-1}(1-\theta)^{\beta-1}, 0<\theta<1$$

That is

$$k(\theta|y) = \frac{\Gamma(\alpha+\beta+n)}{\Gamma(\alpha+y)\Gamma(\beta+n-y)}\theta^{\alpha+y-1}(1-\theta)^{\beta+n-y-1}, 0<\theta<1$$

and $y = 0, 1, \dots, n$. Hence, the posterior pdf is a beta density function with parameters $(\alpha + y, \beta + n - y)$. We take squared error loss, i.e., $\mathcal{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$, as the loss function. Then, the Bayesian point estimate of θ is the mean of this beta pdf which is

$$\delta(y) = \frac{\alpha + y}{\alpha + \beta + n}$$

5.2 Bayesian Interval Estimation

For fixed α, we can find two functions $u(x)$ and $v(x)$ so that the conditional probability

$$P(u(x) < \Theta < v(x) | X = x) = \int_{u(x)}^{v(x)} k(\theta|x)\, d\theta = 1 - \alpha$$

which is defined to be $100(1 - \alpha)\%$ Bayesian interval estimates of θ. This interval is often called **credible interval**, so as not to confuse them with confidence interval.

Example 5.3:

Recall Example 5.1 where $X_1, X_2, \dots, X_n$ is a random sample from a Poisson distribution with mean θ and a $\Gamma(\alpha, \beta)$ prior, with α and β known, is considered. As given, the posterior pdf is a $\Gamma\left(y + \alpha, \frac{\beta}{n\beta+1}\right)$ pdf, where $y = \sum_{i=1}^{n} x_i$, i.e.

$$\theta | X_i \sim \Gamma\left(y + \alpha, \frac{\beta}{n\beta + 1}\right)$$

Find:

a) Bayes' point estimator of Θ using the squared error loss function.
b) Bayes' point estimator of Θ using the absolute error loss function.
c) $(1-\xi)100\%$ credible interval for Θ.
d) a and c when, $\alpha = 2, \beta = 4, n = 12, y = 8, \xi = 0.05$.

Solution:

a) If we use the squared error loss function, the Bayes' point estimate of Θ is the mean of the posterior

$$\delta(y) = \frac{\beta(y+\alpha)}{n\beta+1}$$

b) If we use the absolute error loss function, the Bayes' point estimate of Θ is the median of the posterior or it is the solution, m, of the following equation:

$$\int_0^m \frac{(n\beta+1)^{y+\alpha}\ \theta^{y+\alpha-1} e^{-\frac{(n\beta+1)\theta}{\beta}}}{\Gamma(y+\alpha)\beta^{y+\alpha}} d\theta = \frac{1}{2}$$

c) To obtain a credible interval, from that the posterior distribution of Θ we get that

$$\frac{2(n\beta+1)}{\beta}\Theta \sim \Gamma(y+\alpha, 2) \Leftrightarrow \frac{2(n\beta+1)}{\beta}\Theta \sim \chi^2_{(2(y+\alpha))}$$

Based on this, the following interval is a $(1-\xi)100\%$ credible interval for Θ

$$P\left(\chi^2_{(1-\frac{\xi}{2}, 2(y+\alpha))} < \frac{2(n\beta+1)}{\beta}\Theta < \chi^2_{(\frac{\xi}{2}, 2(y+\alpha))}\right) = 1-\alpha$$

or

$$\Theta \in \left(\frac{\beta}{2(n\beta+1)}\chi^2_{(1-\frac{\xi}{2}, 2(y+\alpha))}, \frac{\beta}{2(n\beta+1)}\chi^2_{(\frac{\xi}{2}, 2(y+\alpha))}\right)$$

where $\chi^2_{(1-\frac{\xi}{2},2(y+\alpha))}$ and $\chi^2_{(\frac{\xi}{2},2(y+\alpha))}$ are the lower and upper χ^2 quantiles for a χ^2 distribution with $2(y+\alpha)$ degrees of freedom.

d) If $\alpha = 2, \beta = 4, n = 12, y = 8, \xi = 0.05,$ then the point estimator is

$$\delta(y) = \frac{4(8+2)}{48+1} = 0.8163$$

and the 95% credible interval for Θ is

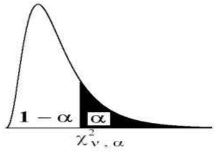
$$\Theta \in \left(\frac{4}{2(48+1)}\chi^2_{(0.975,20)}, \frac{4}{2(48+1)}\chi^2_{(0.025\ ,20)}\right)$$

From χ^2 distribution Table: $\chi^2_{(0.025\ ,20)} = 34.17,\ \ \chi^2_{(0.975,20)} = 9.59$, thus

$$\Theta \in (0.3914, 1.3947)$$

TABLE I

1



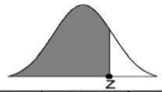
Percentage Points of the χ^2 Distribution; $\chi^2_{\nu,\alpha}$

$P(\chi^2 > \chi^2_{\nu,\alpha}) = \alpha$

	α														
ν	**0.001**	**0.005**	**0.010**	**0.025**	**0.050**	**0.100**	**0.250**	**0.500**	**0.750**	**0.900**	**0.950**	**0.975**	**0.990**	**0.995**	**0.999**
1	10.83	7.88	6.63	5.02	3.84	2.71	1.32	0.45	0.10	0.02					
2	13.82	10.60	9.21	7.38	5.99	4.61	2.77	1.39	0.58	0.21	0.10	0.05	0.02	0.01	
3	16.27	12.84	11.34	9.35	7.81	6.25	4.11	2.37	1.21	0.58	0.35	0.22	0.11	0.07	0.02
4	18.47	14.86	13.28	11.14	9.49	7.78	5.39	3.36	1.92	1.06	0.71	0.48	0.30	0.21	0.09
5	20.52	16.75	15.09	12.83	11.07	9.24	6.63	4.35	2.67	1.61	1.15	0.83	0.55	0.41	0.21
6	22.46	18.55	16.81	14.45	12.59	10.64	7.84	5.35	3.45	2.20	1.64	1.24	0.87	0.68	0.38
7	24.32	20.28	18.48	16.01	14.07	12.02	9.04	6.35	4.25	2.83	2.17	1.69	1.24	0.99	0.60
8	26.12	21.95	20.09	17.53	15.51	13.36	10.22	7.34	5.07	3.49	2.73	2.18	1.65	1.34	0.86
9	27.88	23.59	21.67	19.02	16.92	14.68	11.39	8.34	5.90	4.17	3.33	2.70	2.09	1.73	1.15
10	29.59	25.19	23.21	20.48	18.31	15.99	12.55	9.34	6.74	4.87	3.94	3.25	2.56	2.16	1.48
11	31.26	26.76	24.72	21.92	19.68	17.28	13.70	10.34	7.58	5.58	4.57	3.82	3.05	2.60	1.83
12	32.91	28.30	26.22	23.34	21.03	18.55	14.85	11.34	8.44	6.30	5.23	4.40	3.57	3.07	2.21
13	34.53	29.82	27.69	24.74	22.36	19.81	15.98	12.34	9.30	7.04	5.89	5.01	4.11	3.57	2.62
14	36.12	31.32	29.14	26.12	23.68	21.06	17.12	13.34	10.17	7.79	6.57	5.63	4.66	4.07	3.04
15	37.70	32.80	30.58	27.49	25.00	22.31	18.25	14.34	11.04	8.55	7.26	6.26	5.23	4.60	3.48
16	39.25	34.27	32.00	28.85	26.30	23.54	19.37	15.34	11.91	9.31	7.96	6.91	5.81	5.14	3.94
17	40.79	35.72	33.41	30.19	27.59	24.77	20.49	16.34	12.79	10.09	8.67	7.56	6.41	5.70	4.42
18	42.31	37.16	34.81	31.53	28.87	25.99	21.60	17.34	13.68	10.86	9.39	8.23	7.01	6.26	4.90
19	43.82	38.58	36.19	32.85	30.14	27.20	22.72	18.34	14.56	11.65	10.12	8.91	7.63	6.84	5.41
20	45.31	40.00	37.57	34.17	31.41	28.41	23.83	19.34	15.45	12.44	10.85	9.59	8.26	7.43	5.92
21	46.80	41.40	38.93	35.48	32.67	29.62	24.93	20.34	16.34	13.24	11.59	10.28	8.90	8.03	6.45
22	48.27	42.80	40.29	36.78	33.92	30.81	26.04	21.34	17.24	14.04	12.34	10.98	9.54	8.64	6.98
23	49.73	44.18	41.64	38.08	35.17	32.01	27.14	22.34	18.14	14.85	13.09	11.69	10.20	9.26	7.53
24	51.18	45.56	42.98	39.36	36.42	33.20	28.24	23.34	19.04	15.66	13.85	12.40	10.86	9.89	8.08
25	52.62	46.93	44.31	40.65	37.65	34.38	29.34	24.34	19.94	16.47	14.61	13.12	11.52	10.52	8.65
30	59.70	53.67	50.89	46.98	43.77	40.26	34.80	29.34	24.48	20.60	18.49	16.79	14.95	13.79	11.59
40	73.40	66.77	63.69	59.34	55.76	51.81	45.62	39.34	33.66	29.05	26.51	24.43	22.16	20.71	17.92
50	86.66	79.49	76.15	71.42	67.50	63.17	56.33	49.33	42.94	37.69	34.76	32.36	29.71	27.99	24.67
60	99.61	91.95	88.38	83.30	79.08	74.40	66.98	59.33	52.29	46.46	43.19	40.48	37.48	35.53	31.74
70	112.32	104.21	100.43	95.02	90.53	85.53	77.58	69.33	61.70	55.33	51.74	48.76	45.44	43.28	39.04
80	124.84	116.32	112.33	106.63	101.88	96.58	88.13	79.33	71.14	64.28	60.39	57.15	53.54	51.17	46.52
90	137.21	128.30	124.12	118.14	113.15	107.57	98.65	89.33	80.62	73.29	69.13	65.65	61.75	59.20	54.16
100	149.45	140.17	135.81	129.56	124.34	118.50	109.14	99.33	90.13	82.36	77.93	74.22	70.06	67.33	61.92

TABLE II

Areas Under The Standard Normal Curve

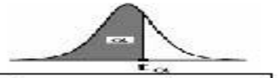


2

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
-3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
-3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
-3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
-3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
-2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	0.0014
-2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
-2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
-2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
-2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
-2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
-2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
-2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
-2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
-2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
-1.4	0.0808	0.0793	0.0778	0.0764	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
-1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
-1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
-1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
-1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
-0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
-0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
-0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
-0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
-0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
-0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
-0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
-0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
-0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
-0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

TABLE III

Critical Values of the t-distribution (t_α)



ν=df	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$	$t_{0.995}$
1	3.078	6.314	12.706	31.821	63.657
2	1.886	2.920	4.303	6.965	9.925
3	1.638	2.353	3.182	4.541	5.841
4	1.533	2.132	2.776	3.747	4.604
5	1.476	2.015	2.571	3.365	4.032
6	1.440	1.943	2.447	3.143	3.707
7	1.415	1.895	2.365	2.998	3.499
8	1.397	1.860	2.306	2.896	3.355
9	1.383	1.833	2.262	2.821	3.250
10	1.372	1.812	2.228	2.764	3.169
11	1.363	1.796	2.201	2.718	3.106
12	1.356	1.782	2.179	2.681	3.055
13	1.350	1.771	2.160	2.650	3.012
14	1.345	1.761	2.145	2.624	2.977
15	1.341	1.753	2.131	2.602	2.947
16	1.337	1.746	2.120	2.583	2.921
17	1.333	1.740	2.110	2.567	2.898
18	1.330	1.734	2.101	2.552	2.878
19	1.328	1.729	2.093	2.539	2.861
20	1.325	1.725	2.086	2.528	2.845
21	1.323	1.721	2.080	2.518	2.831
22	1.321	1.717	2.074	2.508	2.819
23	1.319	1.714	2.069	2.500	2.807
24	1.318	1.711	2.064	2.492	2.797
25	1.316	1.708	2.060	2.485	2.787
26	1.315	1.706	2.056	2.479	2.779
27	1.314	1.703	2.052	2.473	2.771
28	1.313	1.701	2.048	2.467	2.763
29	1.311	1.699	2.045	2.462	2.756
30	1.310	1.697	2.042	2.457	2.750
35	1.3062	1.6896	2.0301	2.4377	2.7238
40	1.3030	1.6840	2.0210	2.4230	2.7040
45	1.3006	1.6794	2.0141	2.4121	2.6896
50	1.2987	1.6759	2.0086	2.4033	2.6778
60	1.2958	1.6706	2.0003	2.3901	2.6603
70	1.2938	1.6669	1.9944	2.3808	2.6479
80	1.2922	1.6641	1.9901	2.3739	2.6387
90	1.2910	1.6620	1.9867	2.3685	2.6316
100	1.2901	1.6602	1.9840	2.3642	2.6259
120	1.2886	1.6577	1.9799	2.3578	2.6174
140	1.2876	1.6558	1.9771	2.3533	2.6114
160	1.2869	1.6544	1.9749	2.3499	2.6069
180	1.2863	1.6534	1.9732	2.3472	2.6034
200	1.2858	1.6525	1.9719	2.3451	2.6006
∞	1.282	1.645	1.960	2.326	2.576