STS 312

STATISTICAL INFERENCE PROCEDURES

LECTURE ONE

STATISTICAL INFERENCE PROCEDURES

INTRODUCTION

Statistics is the study of how information should be employed to reflect on, and give guidance for action in a practical situation involving uncertainty. Any statistical procedure which utilizes information to obtain a description of the practical situation (through a probability model) is an inferential procedure. The study of such procedure will be termed statistical inference. A procedure with the wider aim of suggesting action to be taken in the practical situation, by processing information relevant to that situation, is a decision-making procedure. The study of such procedures is termed statistical decision-making.

A decision problem means the choice between several possible courses of action: this will have observable consequences, which may be used to test its rightness. An inference concerns the degree of belief, which need not have any consequences, though it may. For example, the question "Shall I eat this apple?: is a matter of decision, with possible highly satisfactory or uncomfortable outcomes.

"Is this apple green?" is a question of belief. A statistical inference carries us from observations to conclusions about the population sampled. Statistical inferences involve the data, a specification of the set of possible populations sampled, a question concerning the true populations. The theory of statistical decision deals with the action to take on the basis of statistical information. Decisions are based not only the considerations listed for inferences, but also on an assessment of the losses resulting from wrong decisions, and on prior information, as well as, on a specification of a set of possible decisions.

Point Estimation

Consider a random sample of size n from a population with p.d.f, $f(x, \theta)$. The term random sample may refer either to the set of random variables $X_1, X_2, \ldots, X_n$ or to the observed data $x_1, x_2, \ldots, x_n$.

Definition 1: Statistic

A function of the random sample, $T = t(X_1, X_2, \ldots, X_n)$, that does not depend on any random parameter is called a statistic. A statistic is also a random variable, the distribution of which depends on the distribution of a random sample and on the form of the function $t(X_1, X_2, \ldots, X_n)$. A particular value of the estimator, T is called an estimate.

Loss Function and Risk Function

When an estimate differs from the true value of the parameter being estimated, one may consider the loss involved to be a function of this difference. We shall assume that the loss increases as the square of the difference. In this case, the means square error (MSE) criterion considers the average squared error loss associated with the estimator.

Definition 2: Loss Function

If T is an estimator of $t(\theta)$, then a loss function is any real-valued function $L(t, \theta)$, such that

$$L(t; \theta) \geq 0 \text{ for every } t \tag{1.1}$$

and

$$L(t; \theta) = 0 \text{ when } t = t(\theta) \tag{1.2}$$

Definition 3: Risk Function

The risk function is defined as the expected value of the loss function. That is

$$R_T(\theta) = E[L(T;\, \theta)] \tag{1.3}$$

If a parameter or a function of a parameter is being estimated, one may choose an appropriate loss function depending on the problem, and then try to find an estimator, the average loss (or risk) function that is small for all possible values of the parameter. If the loss function is taken to be squared error, then the risk becomes the MSE. Another reasonable loss function is absolute error, whose risk function is given by

$$R_T(\theta) = E|T - t(\theta)| \tag{1.4}$$

Definition 4: Admissible Estimator

An estimator T_1 is a better estimator than T_2 iff

(i) $R_{T_1}(\theta) \leq R_{T_2}(\theta)$ for all $\theta \in \Omega$ and

(ii) $R_{T_1}(\theta) < R_{T_2}(\theta)$ for at least one θ.

An estimator T is admissible iff there is no better estimator.

LECTURE TWO

DECISION THEORY (DT) APPROACH

In DT, the decision maker chooses an action 'a' from a set of all possible actions based on the observation of a random variable, or data, X, which has a probability distribution that depends on a parameter θ called the state of nature. The set of all possible values of θ is denoted by (H). The decision is made by a statistical decision function d, which maps the sample space (the set of all possible data values) onto the action space A. Denoting the data by X, the action is random and is given as $a = d(X)$.

By taking the action $a = d(X)$, the decision makers incurs a loss, $L(\theta;\ d(X))$, which depends on both θ and $d(X)$. The comparison of different decision functions is based on the risk function, or expected loss,

$$R(\theta;\ d(X)) = E[L(\theta;\ d(X))] \tag{1.1}$$

Here, the expectation is taken with respect to the probability distribution of X, which depends on θ. Note that the risk function depends on the true state of nature, θ, and on the decision function, $d(X)$. Decision theory is concerned with methods of determining "good" decision functions, i.e. decision functions that have small risk

2. Bayes Rule and Minimax Rule

2.1 Minimax Rule (MR)

The MR proceeds as follows: for a given decision function $d(X)$, consider the worst that the risk could be:

$$\max_{\theta \in (H)} [R(\theta) d(X)].$$

Then choose a decision function, d^*, that minimizes this maximum risk

$$\min_{d} \left\{ \max_{\theta \in (H)} \left[R\left(\theta;\ d(X)\right) \right] \right\}$$

Such a decision rule, if it exists, is called a minimax rule.

The weakness of the minimax method is that it is a very conservative procedure. It places all its emphasizes on guarding against the worst possible case. The worst case may not likely occur.

To make this idea more precise, we can assign a probability distribution to the state of nature θ; this distribution is called the prior distribution of θ. Given such a prior distribution, we can calculate the Bayes risk of a decision function *d*:

$$B(d) = E[R(H); d(X)]$$

Here the expectation is taken with respect to the probability distribution of both (H) and X. By the property of iterated conditional expectation, the Bayes risk can be expressed as

$$B(d) = E\left[E\left\{L\left(\theta, d(X)\right) | (H) = \theta\right\}\right]$$

where the inner expectation is conditional on $(H) = \theta$ and the outer expectation is taken with respect to the distribution of (H). The Bayes risk is the average of the risk function with respect to the prior distribution of θ. A function that minimizes the Bayes risk is called a Bayes rule.

Example 1: Consider a loss function and probability distribution below:

	(H)	
A	θ_1	θ_2
a_1	0	400
a_2	100	0

X	θ_1	θ_2
x_1	0.60	0.10
x_2	0.30	0.20
x_3	0.10	0.70

We shall consider the following four decision rules:

d	x_1	x_2	x_3
d_1	a_1	a_1	a_1
d_2	a_1	a_2	a_2
d_3	a_1	a_1	a_2
d_4	a_2	a_2	a_2

To apply the minimax rule, we compute the risk of each of the decision functions in the case where $\theta = \theta_1$ and in the case where $\theta = \theta_2$. For the case $\theta = \theta_1$, each risk function is computed as

$$R\big(\theta_1, d_i(X)\big) = E\big[\big(\theta_1, d_i(X)\big)\big]$$

$$= \sum_{j=1}^{3} L\big(\theta_1, d_i(X)\big) P(X = x; |\theta = \theta_1)$$

We have

$R\big(\theta_1, d_1(X)\big) = 0(0.60) + 0(0.30) + 0(0.10) = 0$

$R\big(\theta_1, d_2(X)\big) = 0(0.60) + 100(0.30) + 100(0.10) = 40$

$R\big(\theta_1, d_3(X)\big) = 0(0.60) + 0(0.30) + 100(0.10) = 10$

$R\big(\theta_1, d_4(X)\big) = 100(0.60) + 100(0.30) + 100(0.10) = 100$

Similarly, for $\theta = \theta_2$, we have

$R\big(\theta_2, d_1(X)\big) = 400$; $R\big(\theta_2, d_2(X)\big) = 40$; $R\big(\theta_2, d_3(X)\big) = 120$; $R\big(\theta_2, d_4(X)\big) = 0$.

To find the minimax rule, we note that the maximum values of d_1, d_2, d_3 and d_4 are 400, 40, 120 and 100, respectively, thus, d_2 is the minimax rule.

Bayes Rule

Suppose we assume a prior distribution $\pi(\theta_1) = 0.80$ and $\pi(\theta_2) = 0.20$. Using this prior distribution and the risk functions computed above, we find for each decision function its Bayes risk,

$$B(d) = E[R(\theta, d(X))]$$

$$= R\left(\theta_1, d(X)\pi(\theta_1) + R\big(\theta_2, d(X)\big)\pi(\theta_2)\right)$$

Thus we have

$$B(d_1) = 0(0.8) + 400(0.2) = 80$$

$$B(d_2) = 40(0.8) + 40(0.2) = 40$$

$$B(d_3) = 10(0.8) + 120(0.2) = 32$$

$$B(d_4) = 100(0.8) + 0(0.2) = 80$$

Given Bayes rule = d_3

Given $f(x;\theta) = \frac{e^{-\theta}\theta^x}{x!}, \ x = 0,1,\ldots$

$= 0$

(i) Calculate the risk function if

$L(\theta; d(X)) = (d(X) - \theta)/\theta$ for $d(x) = cx$

(ii) Calculate the risk if c = ₦11 and $\theta = 0.10$

(iii) Determine the value of c for which the risk function is minimum.

$$\frac{(d(X) - \theta)^2}{\theta} = (cx - \theta)^2/\theta$$

$$= \frac{1}{\theta}(c^2x^2 - 2\theta cx + \theta^2)$$

$$= \left(\frac{c^2x^2}{\theta} - 2cx + \theta\right)$$

$$R(\theta; d(X)) = E[L(\theta; d(X))]$$

$$= E\left[\frac{c^2x^2}{\theta} - 2cx + \theta\right]$$

$$= c^2\frac{(\lambda^2+\lambda)}{\theta} - 2c\lambda + \theta$$

$$= c^2\left(\frac{\theta^2+\theta}{\theta}\right) - 2c\theta + \theta$$

$$= c^2(\theta + 1) - 2c\theta + \theta$$

$$= c^2 + \theta(c^2 - 2c + 1)$$

$$= c^2 + \theta(c - 1)^2.$$

Example 1

Given $f(x/\theta) = \frac{e^{-\theta}\theta^x}{x!}, \ x = 0,1,\ldots$

Calculate $L(\theta, d) = (d - \theta)^2$ for $d(X) = CX$

$$= CX \text{ so that } \frac{(CX - \theta)^2}{\theta} = \frac{C^2X^2 - 2CX\theta - \theta^2}{\theta}$$

$$R(\theta, d) = \sum_{x=0}^{\infty} X^2 f(X/u) = \lambda^2 + \lambda$$

$$R(\theta, d) = \frac{1}{\theta}\{C^2(\lambda^2 + \lambda) - 2C\theta\lambda + \theta^2\}$$

$$= \frac{1}{\theta}\{C^2\theta^2 + C^2\theta - 2C\theta + \theta^2\}$$

$$= C^2\theta + C^2 - 2C\theta + \theta$$

$$= C^2 + C^2\theta - 2C\theta + \theta$$

$$= C^2 + \theta(C^2 - 2C + 1)$$

$$= C^2 + \theta(C - 1)^2$$

Example 2

Given $f(x/\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}$, $x = 0,1,2;\ \ 0 < \theta < 1$

$$= \binom{n}{x}p^x q^{n-x}$$

and $L(\theta, d) = (a - \theta)^2$,

calculate $R(\theta, d)$ for (i) $d(X) = \frac{x}{2} = \frac{\theta(1-\theta)}{2} = \frac{pq}{2}$

(ii) $d(X) = \left(\frac{x+1}{4}\right) = \frac{(2\theta^2 - 2\theta + 1)}{16}$

Find:

(a) $R(\theta, d(X))$ when $d(X) = x/2$

(b) The result in (a) if $p = q = \frac{1}{2}$.

$$f(x/\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}$$

$$L(\theta, a) = \left(\frac{X}{2} - \theta\right)^2 = \frac{X^2}{2} - \theta X + \theta^2$$

$$\sum_{0}^{\infty} \tilde{x} f\,(x/\theta) = n(n-1)p^2 + np = n(n-1)\theta^2 + n\theta$$

$$R(\theta, d) = \frac{1}{4}[n(n-1)\theta^2 + n\theta] - \theta X + \theta^2$$

$$= \frac{1}{4}[n^2\theta^2 - n\theta^2 + n\theta] - \theta X + \theta^2$$

$$= \theta^2 - \frac{\theta^2}{2} + \frac{\theta}{2} - 2\theta^2 + \theta^2$$

$$= \frac{\theta}{2} - \frac{\theta^2}{2}$$

$$= \frac{\theta}{2}(1-\theta).$$

Example 3

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

$$L(\theta, d) = (d-\theta)^2/\theta$$

Find $R(\theta, d)$ for $d(X) = CX$

Solution:

$$\frac{(CX-\theta)^2}{\theta} = \frac{(C^2x^2 - 2Cx\theta - \theta^2)}{\theta}$$

$$R(\theta, d) = \frac{1}{\theta}\sum_X (C^2x^2 - 2Cx\theta - \theta^2)e^{-\theta}\frac{\theta^x}{x!}$$

$$= \frac{1}{\theta}\{C^2(\lambda^2+\lambda) - 2C\theta\lambda + \theta^2\}$$

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

$$= C^2\theta + C^2 - 2C\theta + \theta$$

$$= C^2 + C^2\theta - 2C\theta + \theta$$

$$= C^2 + \theta(C^2 - 2C + 1)$$

$$= C^2 + \theta(C-1)^2$$

EXERCISES

(a) Given $f(x/\theta) = \frac{e^{-\theta}\theta^x}{x!}$, $x = 0,1,\dots$, $\theta > 0$

and $L(\theta, a) = (a-\theta)^2$,

(i) Calculate $R(\theta, d)$ for $d(X) = x$

(ii) Calculate $R(\theta, d)$ for $d(X) = CX$

$L(\theta, a) = (CX - \theta)^2$

$$= C^2\left(X - \frac{\theta}{c}\right)^2$$

$$= C^2\left(X - \theta + \theta\left(1 - \frac{1}{c}\right)^2\right)$$

$$R(\theta, d) = \sum_{i=0}^{\infty} C^2\left(X - \theta + \theta\left(1 - \frac{1}{c}\right)^2\right)\frac{e^{-\theta}\theta^x}{x!}$$

$$= C^2\sum_{x=0}^{\infty} x\frac{e^{-\theta}\theta^x}{x!} - \theta + \theta\left(1 - \frac{1}{c}\right)$$

(iii) $f(x/\theta) = \binom{2}{x}\theta^x(1-\theta)^{2-x}$, $X = 0,1,2$, $0 < \theta < 1$

$L(\theta, a) = (a-\theta)^2$

Calculate $R(\theta, d)$ for $d(X) = \frac{x}{2}$.

Q5

$$L(\theta, a) = \left(\frac{x}{n} - \theta\right)^2 = \frac{x^2}{n^2} - \frac{2x\theta}{n} + \theta^2$$

$$\sum_{0}^{\infty} \tilde{x}\theta^x(1-\theta)^{n-x} = n(n-1)p^2 + np$$

$$= \frac{1}{n^2}(n^2p^2 - up^2 + np)$$

$$= \left(\theta^2 - \frac{\theta^2}{n} + \frac{\theta}{n}\right) - 2\theta^2 + \theta^2$$

$$= \frac{\theta}{n} - \frac{\theta^2}{n}$$

$$= \frac{\theta}{n}(1-\theta).$$

Q7

$$f\left(\frac{x}{\theta}\right) = \frac{e^{-\theta}\theta^{x}}{x!}, \qquad x = 0,1,\ldots,$$

$$L(\theta, a) = (a - \theta)^2,$$

Calculate $R(\theta, d)$ for $d(X) = x$

$$L(\theta, a) = (x - \theta)^2 = x^2 - 2x\theta + \theta^2$$

$$\sum_{x=0}^{\infty} x^2 f(x/\theta) = \lambda^2 + \lambda = \theta^2 + \theta$$

$$\therefore \quad \sum_{x=0}^{\infty} (x - \theta)^2 f(x/\theta) = \theta^2 + \theta - 2\theta^2 + \theta^2$$

$$= \theta.$$

LECTURE THREE

SOME UNIVARIATE DISTRIBUTIONS

1.0 Binomial Distribution

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \ x = 0,1,2,\ldots,n \qquad (1.1)$$

$$= 0, \text{ otherwise}$$

$$E(X) = \sum_{x=0}^{n} xf(x) = \binom{n}{x} x \binom{n}{x} p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} \frac{n(n-1)!}{x(x-1)!\,(n-x)!} x\,p\,.\ p^{x-1}(1-p)^{n-x}$$

$$= up \sum \frac{(n-1)!}{(x-1)!\,(n-x)!} p^{x-1}(1-p)^{n-x}$$

$$= up \sum_{t}^{n-1} \binom{n-1}{t} p^t (1-p)^{n-1-t}$$

$$= up(1) = up \qquad (1.2)$$

$$E(X^2) = \sum_{x=0}^{n} \binom{n}{x} x^2\, p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} [x(x-1) + x]\, p^x (1-p)^{n-x}$$

$$= \sum_{x=0}^{n} x(x-1) \binom{n}{x} p^x q^{n-x} + np, \qquad \text{using (1.2)}$$

$$= \sum_{x=0}^{n} x(x-1) \frac{n(n-1)(n-2)!}{x(x-1)(x-2)!\,(n-x)!} p^2 p^{x-2} q^{n-x} + np$$

$$= n(n-1)p^2 \sum \frac{(n-2)!}{(x-2)!\,(n-x)!} p^{x-2} q^{n-x} + np$$

$$= n(n-1)p^2 \sum_{t} \binom{n-2}{t} p^t q^{n-2-t} + np$$

$$E(X^2) = n(n-1)p^2(1) + up \qquad (1.3)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$Var(X) = n(n-1)p^2 + np - [E(X)]^2$$

$$= n^2p^2 - np^2 + np - n^2p^2, \text{ using (1.2) \& (1.3)}$$

$$= np - np^2$$

$$= np(1-p)$$

$$= npq \tag{1.4}$$

2.0 POISSON DISTRIBUTION

$$P(X) = e^{-\lambda}\lambda^x | x!\,, \quad x = 0,1,2,\ldots \tag{2.1}$$

$= 0$, otherwise

$$E(X) = \sum_{x=0}^{\infty}\left[xe^{-\lambda}\lambda^x | x!\right]$$

$$= \sum_{x=0}^{\infty}\left[xe^{-\lambda}\lambda.\lambda^{x-1} | x(x-1)!\right]$$

$$= e^{-\lambda}\lambda \sum (\lambda^x - 1) | (x-1)!$$

$$= e^{-\lambda}\lambda \sum_{t}^{\infty}\left(\frac{\lambda^t}{t!}\right), \quad t = (x-1)$$

$$= e^{-\lambda}\lambda e^{\lambda}$$

$$= \lambda \tag{2.2}$$

$$E(X^2) = \sum_{x=0}^{\infty}\left[x^2e^{-\lambda}\lambda^x | x!\right]$$

$$= \sum_{x=0}^{\infty}\left\{[x(x-1) + x] \quad e^{-\lambda}\lambda^x | x!\right\}$$

$$= \sum_{x=0}^{\infty}\left\{x(x-1)e^{-\lambda}\lambda^x | x!\right\} + \lambda, \quad \text{using (2.2)}$$

$$= \sum_{x=0}^{\infty} \{x(x-1)e^{-\lambda}\lambda^2\lambda^{x-2} | x(x-1)(x-2)!\} + \lambda$$

$$= e^{-\lambda}\lambda^2 \sum \frac{\lambda^{x-2}}{(x-2)!} + \lambda$$

$$= e^{-\lambda}\lambda^2 \sum_{t}^{\infty} \frac{\lambda^t}{t!} + \lambda$$

$$= e^{-\lambda}\lambda^2 e^{\lambda} + \lambda$$

$$= \lambda^2 + \lambda \qquad (2.3)$$

$$Var(X) = E(X^2) - [E(X)]^2$$

$$= \lambda^2 + \lambda - (\lambda)^2$$

$$= \lambda^2 + \lambda - \lambda^2$$

$$= \lambda \qquad (2.4)$$

Example:

Given $f(x) = e^{-\theta}\theta^x / x!$, $x = 0,1,\ldots,$ and $L(\theta, a) = (d-\theta)^2/\theta$

(i) Calculate the risk, $R(\theta, d)$ for $d(X) = CX$

(ii) Determine the value of C for which $R(\theta, d)$ is a minimum

(iii) The minimum value of the risk.

Solution

$$\frac{1}{\theta}(Cx-\theta)^2 = \frac{1}{\theta}(C^2x^2 - 2x\theta C + \theta^2)$$

$$R(\theta, d) = E\left[\frac{1}{\theta}(Cx-\theta)^2\right]$$

$$R(\theta, d) = \frac{1}{\theta}\sum_{x=0}^{\infty}\left[(C^2x^2 - 2C\theta x + \theta^2)\frac{e^{-\theta}\theta^x}{x!}\right]$$

$$= \frac{1}{\theta}[C^2(\lambda^2 + \lambda) - 2C\theta(\lambda) + \theta^2(1)]$$

$$= \frac{1}{\theta}[C^2(\theta^2 + \theta) - 2C\theta(\theta) + \theta^2]$$

$$= \frac{1}{\theta}[C^2\theta^2 + C^2\theta - 2C\theta^2 + \theta^2]$$

$$= [C^2\theta + C^2 - 2C\theta + \theta]$$

$$= C^2 + C^2\theta - 2C\theta + \theta$$

$$= C^2 + [\theta(C^2 - 2C + 1)]$$

$$R(\theta, d) = C^2 + \theta(C - 1)^2$$

(ii) The $R(\theta, d)$ is at a minimum when $C = 1$

(iii) At $C = 1$, $R(\theta, d) = 1$.

LECTURE FOUR

POINT ESTIMATION

Let a random variable X have a p.d.f. which is of known functional form but the p.d.f. depends on an unknown parameter θ that may have any value in the set Ω. That is, $f(x;\theta)$, $\theta \in \Omega$ is the p.d.f. of X, where Ω is the parameter space.

Definition 1:

Any statistic whose mathematical expectation is equal to a parameter θ is called an unbiased statistic for the parameter θ. Otherwise the statistic is said to be biased.

Definition 2:

For a given positive integer n, $Y_1 = t(X_1, \ldots, X_n)$ will be called a "best statistic" for a parameter θ if Y_1 is unbiased, $E(Y_1) = \theta$, and if the variance of Y_1 is less than or equal to the variance of every other unbiased statistic for θ.

Example 1:

(i) Show that $\bar{X}$ of a random sample of size n from a distribution having p.d.f. $f(x;\theta) = \frac{1}{\theta}e^{-(x/\theta)}, \begin{matrix} 0 < x < \infty \\ 0 < \theta < \infty \end{matrix};$ is unbiased for θ.

(ii) Compute the variance of $\hat{\theta}_1$

Solution:

$$f(x;\theta) = \frac{1}{\theta}e^{-(x/\theta)}, \qquad \begin{matrix} 0 < x < \infty \\ 0 < \theta < \infty \end{matrix}$$

Set $n = \frac{x}{\theta} \Longrightarrow \begin{matrix} x = \theta u \\ dx = \theta du \end{matrix}$

i.e.

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{1}^{n} x_i\right) = \frac{1}{n}\sum_{1}^{n}\int_0^\infty \frac{x}{\theta}e^{-(x/\theta)}dx \qquad (1.1)$$

$$\int_0^\infty \frac{x}{\theta}e^{-(x/\theta)}dx = \theta\int_0^\infty ue^{-u}.du = \theta\Gamma(2)$$

$$= \theta.1 = \theta \qquad (\Gamma(n) = (n-1)!) \qquad (1.2)$$

Using (1.2) in (1.1) we have

$$E(\bar{X}) = \frac{1}{n}\sum_{1}^{h} \theta = \theta \qquad (1.2.1)$$

$$E(\bar{X}^2) = E\left(\frac{1}{n}\sum_{1}^{n} x_i\right)^2 = \frac{E}{n^2}\left(\sum_{1}^{h} x_i^2 + \sum_{i=j}\sum x_i x_j\right) \qquad (1.3)$$

$$E(x^2) = \int_0^\infty \frac{x^2}{\theta} e^{-\left(\frac{x}{\theta}\right)} dx$$

$$= \int_0^\infty \frac{\theta^2 u^2}{\theta} e^{-u} \,.\, \theta du$$

$$= \theta^2 \int_0^\infty u^2\, e^{-u}\, du = \theta^2 \Gamma(3)$$

$$= 2\theta^2 \qquad (1.4)$$

$$E(x_i x_j) = \theta^2, \text{ from } (1.2)$$

From (1.2)

$$E(\bar{X}^2) = \frac{1}{n^2}\left(\sum_{1}^{h} 2\,\theta^2 + \sum_{i=j}\sum \theta^2\right)$$

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

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

$$= \frac{2\theta^2}{n} + \theta^2 - \frac{\theta^2}{n}$$

$$E(\bar{X}^2)$$

$$V(\hat{\theta}) = E(\bar{X}^2) - [E(\bar{X})]^2$$

$$= \frac{\theta^2}{n} + \theta^2 - \theta^2$$

$$= \frac{\theta^2}{n}$$

Exercise 1:

Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the uniform distribution

$$f(x;\theta) = \frac{1}{\theta}, \quad 0 < x < \theta; \quad 0 < \theta < \infty$$
$$= 0, \qquad \text{otherwise}$$

(i) Show that $4Y_1$ is an unbiased statistic for θ.

(ii) Compute the variance for $\hat{\theta}$.

1. Let Y_1 and Y_2 be two statistically independent unbiased statistic for θ. Say the variance of Y_1 is twice the variance of Y_2. Find the constants k_1 and k_2 so that $k_1Y_1 + k_2Y_2$ is an unbiased statistic with smallest possible variance for such a linear combination.

Q2

$$E(k_1Y_1 + k_2Y_2) = \theta$$

$$=> k_1Y_1 + k_2Y_2 = \hat{\theta}$$

$$\text{i e. } k_1\theta + k_2\theta = \theta$$

$$\text{or } k_1 + k_2 = 1 \qquad (1.1)$$

$$V(\hat{\theta}) = V(k_1Y_1 + k_2Y_2)$$

$V(Y_1) = 2V(Y_2)$; let $V(Y_1) = 2\sigma$ so that $V(Y_2) = \sigma$

We have

$V(\hat{\theta}) = 2k_1^2\sigma + k_2^2\sigma = 2k_1^2\sigma + (1 - k_1)^2\sigma$, from (1.1)

$= 2k_1^2\sigma + (1 - 2k_1 + k_1^2)\sigma$

$V(\hat{\theta}) = 3k_1^2\sigma - 2k_1\sigma + \sigma$

$\frac{\partial V(\hat{\theta})}{\partial k} = 0 \Longrightarrow 6k_1\sigma - 2\sigma = 0$

i.e. $(6k_1 - 2)\sigma = 0$

since $\sigma \neq 0$,

$$6k_1 - 2 = 0 \text{ or } \left.\begin{aligned} k_1 &= \frac{1}{3} \\ \text{and } k_2 &= \frac{2}{3} \end{aligned}\right\}$$

Definition 1

If T is an estimator of $t(\theta)$

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

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

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

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

$= Var(T) + [B(T)]^2$

$= Var(T)$, if $E(T) = t(\theta)$

Definition 2:

Let $t(\theta)$ be estimable. An estimator $T = t(X_1, \dots, X_n)$ is said to be a VMVU estimator of $t(\theta)$ if it is unbiased and has the smallest variance within the class of all unbiased estimators of $t(\theta)$ under all $\sigma \in \Omega$.

In many cases of interest, a VMVU estimator does exist. The problem is how one would go about searching for it (if it exists). There are two approaches which may be used. The first is appropriate when complete sufficient statistics are available and provides us with a UMVU estimator. The second approach is to first determine a lower bound for the variances of all estimators in the class under classification and then try to determine an estimator whose variance is equal to this lower bound. The Cramer-Rao inequality is instrumental to this approach.

1. Method of Estimation

1.1 Method of Maximum Likelihood

We had earlier discussed the method of least squares. According to the principle of maximum likelihood, we should choose the estimator which makes the likelihood function a maximum. That is, t_n will be the maximum likelihood estimator (m.l.e.) if

$$L(x_1, x_2, \dots, x_n, t_n), \quad L(x_1, x_2, \dots, x_n, t'_n)$$

for any other estimator t'_n. If L is a differentiable function of θ then $t_n(x_1, x_2, \dots, x_n)$ is the solution (if any) of

$$\frac{\delta L}{\delta\theta} = 0 \quad \text{with} \quad \frac{\delta^L L}{\delta\theta^2} < 0$$

Since L is positive, the first equation is equivalent to

$$\frac{\delta \log L}{\delta\theta} = 0,$$

a form which is more convenient in practice.

Example: For a random sample from a normal population, find the m.l.e. for the population mean, when the variance is known.

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

$$L = (2\pi\sigma^2)^{-\frac{n}{2}} e^{-\sum_1^h (x_i-\mu)^2} / 2\sigma^2$$

$$\log L = -\frac{n}{2}\log 2\pi\sigma^2 - \sum_1^h (x_i - \mu)^2 \Big/ 2\sigma^2$$

$$\frac{\delta \log L}{\delta\mu} = 0 \Rightarrow \sum_1^h (x_i - \mu) \Big/ \sigma^2 = 0$$

$$\Rightarrow \sum_1^h x_i - n\mu = 0$$

$$\text{or} \quad \hat{\mu} = \frac{1}{n}\sum_{i=1}^h x_i$$

$$= \bar{x}$$

Properties of Maximum Likelihood Estimators

(a) Maximum likelihood estimators are consistent if

(i) the density function $f(x; \theta)$ is continuous in x throughout its range and if

(ii) $f(x; \theta)$ is continuous and monotonic in θ in some θ interval containing the true value θ_0 and for all x, in some x-interval, then the m.l.e., $\hat{\theta}$ is consistent.

(b) The distribution of m.l.e. tends to normality for large samples. More specifically, if

(i) $f(x;\theta)$ is continuous in x throughout its range, and if

(ii) in a θ-interval containing the true value θ_0, $\frac{\partial f}{\partial \theta}$ is continuous in θ for every x, $x^2 \frac{\partial f}{\partial \theta}$ approaches continuous function of θ as x tends to infinity, and $\frac{\partial f}{\partial \theta}$ does not vanish in some interval, then for large n, the m.l.e. of θ will tend to be normally distributed with variance given by

$$\frac{1}{Var(\hat{\theta})} = \int_{-\infty}^{\infty} \frac{1}{f}\left(\frac{\partial f}{\partial \theta}\right)^2 dx = n\left(\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta} \log f\right)^2 dx$$

$$= nE\left(\frac{\partial}{\partial \theta} \log f\right)^2$$

If the range is independent of f, or if f and $\frac{\partial f}{\partial \theta}$ vanish at the extremity of the range which depends on θ, we have the alternative form, namely,

$$\frac{1}{Var(\hat{\theta})} = -nE\left(\frac{\partial^2}{\partial \theta^2} \log f\right)$$

(c) Maximum likelihood estimators are not efficient. That is, in the cl.ass of estimators which for large n tends to be normally distributed about population parameters as mean, the variance of the m.l.e. will be less than or equal to that of any other estimator. That is, if t is any other such estimator,

$$Var(t) \geq Var(\hat{\theta}).$$

(d) Maximum likelihood estimators are sufficient, if sufficient estimators exist. That is, if a sufficient estimator exists, it is a function of the m.l.e.

(e) Maximum likelihood estimators have the invariance property. That is, if $\hat{\theta}$ is a m.l.e. for θ, then $f(\hat{\theta})$ will be a m.l.e. for $f(\theta)$.

(f) Maximum likelihood estimators are not necessarily unbiased.

Example 2: Using example, find the efficiency of \$\bar{X}$ based on a random sample of size.

The efficiency of an unbiased estimator $\bar{X}$ is given by

$$e(\bar{X}) = \frac{1}{Var(X)E\left[\frac{\log f(X;\theta)}{\partial\theta}\right]^2}$$

$$\frac{\delta}{\partial\theta}\log f(X;\theta) = \frac{\delta}{\partial\mu}\log(X;\mu,\sigma^2) = \frac{x-\mu}{\sigma^2}$$

$$\left[\frac{\delta}{\partial\mu}\log(X;\mu,\sigma^2)\right]^2 = \left(\frac{x-\mu}{\sigma^2}\right)^2$$

$$E\left[\frac{\delta}{\partial\mu}\log(X;\mu,\sigma^2)\right]^2 = \frac{1}{\sigma^4}E(x-\mu)^2 = \frac{1}{\sigma^4}(\sigma^2)$$

$$= \frac{1}{\sigma^2}$$

$$Var(\bar{X}) = V\left(\frac{1}{n}\sum_{i=1}^{h} x_i\right) = \frac{1}{n^2}\sum_{i=1}^{h} Var(x_i) = \frac{1}{n^2}\sum_{i=1}^{h}\sigma^2$$

$$= \frac{\sigma^2}{n}$$

$$e(\bar{X}) = \frac{1}{\frac{\sigma^2}{n}n(1/\sigma^2)} = 1$$

showing that $\bar{X}$ is efficient.

Exercise

Let $X_1, X_2, \ldots, X_n$ represent a random sample from

$$f(x;\theta) = \frac{1}{\theta}e^{-\frac{x}{\theta}}, \qquad 0 < x < \infty, \ 0 < \theta < \infty$$

$$= 0, \ \text{elsewhere}$$

Find the m.l.e. of θ.

Method of Moments

The method due to K. Pearson is used in fitting distributions specially of the Pearson type. According to this method, to choose *m* parameters of a population, we equate the first *m* moments of the sample values to the first *m* moments of the population and solve from these *m* equations for the *m* estimators.

Example 3: Use the method of moments to obtain the parameters of the gamma density

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

$$E(X) = \mu = M_1' = \int_0^\infty x \,.\, \frac{x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)}\, dx = \beta\alpha \tag{1.1}$$

$$E(X^2) = M_2' = \int_0^\infty \frac{x^2 x^{\alpha-1}e^{-x/\beta}}{\beta^\alpha\Gamma(\alpha)} = \alpha \tag{1.2}$$

So that $\sigma^2 = M_2' - (M')^2 = \alpha(\alpha+1)\beta^2 - \alpha^2\beta^2$

$$\sigma^2 = \beta^2\alpha \tag{1.3}$$

From (1.1) we have

$$\bar{x} = \beta\alpha \tag{1.4}$$

From (1.3), $s^2 = \beta^2\alpha$ (1.5)

The solution of equations (1.4) and (1.5) gives

$\hat{\alpha} = \bar{x}^2/s^2$

$\hat{\beta} = s^2/\bar{x}$

Moment Method of Estimation

Definition 1: A sample Moment

Let $X_1, X_2, \dots, X_n$ denote a random sample from the density $f(\,.\,)$. The rth sample moment about zero, denoted by M_r', is defined by

$$M_r' = \frac{1}{n}\sum_{i=1}^{n} X_i^r \tag{1.1}$$

In particular, if $r = 1$, we have the sample mean given by

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{1.2}$$

The r^{th} sample moment about the mean $(\bar{X}_n)$, denoted by M_r, is given by

$$M_r = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X}_n)^r \tag{1.3}$$

It is well known that the sample moments reflect the population moments in the sense that the expected value of a sample (about zero) equals the corresponding population moment. Also, the variance of a sample moment is $\left(\frac{1}{n}\right)$ times some function of the population moments. Thus a sample moment can be used to estimate its corresponding population moment (provided the population moment exists).

Let $X_1, X_2, \dots, X_n$ denote a random sample from a population with a density $f(\,.\,)$. The expected value of the r^th sample moment (about zero) is equal to the r^th population moment. That is,

$$E(M_r') = \mu_r' \tag{1.4}$$

(if u_r exists).

For example, the two parameters μ and σ^2 of a normal distribution are moments of the distribution. Therefore they would be estimated by the sample mean $\bar{X}_n$ and sample variance S_n^2.

If a distribution has only one unknown parameter but that parameter is not a moment of the distribution, the parameter may still be estimated by the method of moments by calculating the first moment of the distribution, which will be a function of the parameter, and equating it to $\bar{X}_n$. The solution of the resulting equation for the unknown parameter value will be the desired estimate. Similarly, if the distribution had two unknown parameters that were not moments, the same procedure would be followed with respect to the first two moments of the distribution.

For an illustration for which the parameters are not moments, consider estimating the two parameters of the gamma density with the method of moments. Let the two parameter gamma be given by

$$f(x;\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, 0 < x < \infty$$

$$= 0, \quad \text{otherwise.} \tag{2.1}$$

The population mean, μ is given by

$$E(X) = \int_0^{\infty} \frac{x}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} dx = \bar{x}_n \tag{2.1.1}$$

$$= \int_0^\infty \frac{x^\alpha e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} dx \qquad \text{set} \begin{array}{c} u = \frac{x}{\beta} \\ dx = \beta du \end{array}$$

$$= \int_0^\infty \frac{(u\beta)^\alpha}{\Gamma(\alpha)\beta^\alpha} . \beta e^{-u} du$$

$$= \frac{\beta}{\Gamma(\alpha)} \int_0^\infty u^\alpha e^{-u} du$$

$$= \frac{\beta}{\Gamma(\alpha)} \Gamma(\alpha + 1)$$

$$= \frac{\beta}{(\alpha+1)!} (\alpha)! = \frac{\alpha\beta(\alpha+1)!}{(\alpha+1)!}$$

$$= \alpha\beta = \bar{x}_n \tag{1.2}$$

The second moment is given by

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

$$= \int_0^\infty \frac{x^{\alpha+1} e^{-x/\beta}}{\Gamma(\alpha)\beta^\alpha} dx$$

$$= \int_0^\infty \frac{(u\beta)^{\alpha+1} e^{-u}}{\Gamma(\alpha)\beta^\alpha} . \beta\, du$$

$$= \frac{\beta^2}{\Gamma(\alpha)} \int_0^\infty u^{\alpha+1} e^{-u} du = \frac{\beta^2}{\Gamma(\alpha)} \Gamma(\alpha + 2)$$

$$= \frac{\beta^2}{(\alpha-1)!} (\alpha + 1)!$$

$$= \frac{\beta^2(\alpha+1)\alpha(\alpha-1)!}{(\alpha-1)!}$$

$$= \alpha(\alpha + 1)\beta^2 \tag{1.3}$$

Population variance, $\sigma^2 = E(X^2) - [E(X)]^2 = s_n^2$ (1.4)

i.e. $\sigma^2 = \alpha(\alpha + 1)\beta^2 - (\alpha\beta)^2$

$$= \alpha^2\beta^2 + \alpha\beta^2 - \alpha^2\beta^2 = \alpha\beta^2 = s_n^2 \tag{1.5}$$

From (1.2) and (1.5), we have

$\alpha\beta = \bar{x}_n$ (1.2)

$\alpha\beta^2 = s_n^2$ (1.5)

Dividing (1.5) by (1.2) we have

$$\hat{\beta} = s_n^2 / \bar{x}_n \tag{1.6}$$

Substituting from (1.6) in (1.2), we have

$\alpha(s_n^2/\bar{x}) = \bar{x}_n$ or

$$\hat{\alpha} = \bar{x}^2 / s_n^2 \tag{1.7}$$

Exercise

1. Given $f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2}$, show that $\frac{1}{n}\sum_{i=1}^{n} X_i^2$, where $X_1, X_2, \ldots, X_n$ is a random sample from $f(x;\theta)$ is an unbiased estimator of θ.

2. Given $f(x;\theta) = \frac{X^{\theta-1}e^{-x}}{\Gamma(\theta)}$, $X > 0, \theta > 0$, find a value of c such that CX will be an unbiased estimator of θ.

3. Find the lower bound of the variance for an unbiased estimator of the parameter θ for the Cauchy density

$$f(x;\theta) = \frac{1}{\pi[1+(x-\theta)^2]}$$

$$= 0, \text{ otherwise.}$$

Questions (Point Estimation)

Given $f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{x^2}{2\theta}}, -\infty < x < \infty, \; \theta > 0$. Show that $\frac{1}{n}\sum_{i=1}^{n} X_i^2$, where $X_1, X_2, \ldots, X_n$ is a random sample from $f(x;\theta)$ is an unbiased estimator of θ.

Solution:

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \int_{-\infty}^{\infty} \frac{1}{n}\sum_{1}^{n} x^2 (2\pi\theta)^{-\frac{1}{2}} e^{-\frac{x^2}{2\theta}} dx$$

$$= (2\pi\theta)^{-\frac{1}{2}} \frac{1}{n}\sum_{1}^{n} \int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\theta}} dx \tag{1.1}$$

$$\int_{-\infty}^{\infty} x^2 e^{-\frac{x^2}{2\theta}} dx = 2\int_0^{\infty} x^2 e^{-\frac{x^2}{2\theta}} dx \quad \text{set } u = x^2/2\theta, \qquad \theta du = xdx$$

$$= 2\int_0^{\infty} x \,.\, e^{-\frac{x^2}{2\theta}} \,.\, xdx$$

$$= 2\int_0^{\infty} (2u\theta)^{1/2}\, e^{-u} \,.\, \theta du$$

$$= 2\sqrt{2\theta}\int_0^{\infty} u^{\frac{1}{2}}\ e^{-u} du = 2\sqrt{2\theta}\int_0^{\infty} u^{\frac{3}{2}}\ e^{-1-u} du$$

$$= 2\theta\sqrt{2\theta}\,\Gamma\left(\frac{3}{2}\right)$$

$$= 2\theta\sqrt{2\theta}\,.\frac{1}{2}\Gamma\left(\frac{1}{2}\right), \quad \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$$

$$= \theta\sqrt{2\theta}\,\Gamma\left(\frac{1}{2}\right)$$

$$= \theta\sqrt{2\theta}\ .\ \sqrt{\pi}, \quad \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} \tag{1.2}$$

$$= \theta\sqrt{2\pi\theta}$$

Substituting from (1.2) in (1.1) we have

$$E\left(\frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = \frac{1}{\sqrt{2\pi\theta}}\frac{1}{n}\sum_{i=1}\theta\,\frac{\sqrt{2\pi\theta}}{1} = \frac{1}{n}\sum_{i=1}^{n}\theta$$

$$= \frac{1}{n}n\theta$$

$$= \theta.$$

We have

$C\theta = \theta$

or $C = 1$

Note:

1. $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$
2. $\Gamma(\alpha + 1) = \alpha\Gamma(\alpha)$
3. $\Gamma(\alpha + 2) = \alpha\Gamma(\alpha + 1)$
4. $\Gamma(\alpha) = \frac{\Gamma(\alpha+1)}{\alpha}$

LECTURE FIVE

CONFIDENCE INTERVALS

Definition 1:

Let $X_1, \ldots, X_n$ be a random sample from the density $f(x;\theta)$. Let $T_1 = t_1(X_1, \ldots, X_n)$ and $T_2 = t_2(X_1, \ldots, X_n)$ be two statistics satisfying $T_1 \leq T_2$ for which $P_\theta[T_1 < t(\theta) < T_2] = \gamma$, where γ does not depend on θ. Then the random interval (T_1, T_2) is called a 100 γ percent confidence interval for $\tau(\theta)$; γ is called the confidence coefficient; and T_1 and T_2 are called the lower and upper confidence limits, respectively, for $\tau(\theta)$. A value (t_1, t_2) of the random interval (T_1, T_2) is also called a 100 γ percent confidence interval for $\tau(\theta)$.

Definition 2:

An interval is said to be random if at least one of its end points t_1, t_2 is a random variable.

Example 1:

Let X be $\chi^2_{(16)}$. What is the probability that the random interval $(X, 3.3X)$ contains the point x = 26.3? Compute the expected length of the interval.

Solution

We have $X < 26.3 < 3.3X$ when $X < 26.3$ and $3.3X > 26.3$ or $X > \frac{26.3}{3.3} = 7.96969697 = 7.97$

That is $Prob(7.97 < X < 26.3)$

$= P(X \leq 26.3) - P(X \leq 7.97)$

$= 0.95 - 0.050$

$= 0.90$

The length of the interval is $3.3X - X = 2.3X$.

The expected length is $2.3E(X) = 2.3(16)$

Since X is $\chi^2_{(16)}$, that is 36.80.

Example 2:

Let the random variable X have the p.d.f. $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Compute the probability that the random interval (X, $3X$) includes the point $x = 3$. What is the expected value of the length of this random interval?

Solution:

We have $X < 3$ and $X > 3/3 = 1$. That is we desire $Prob(1 < X < 3)$

$= \int_1^3 e^{-x}dx = [-e^{-x}]_1^3 = [e^{-x}]_1^3$

$= e^{-1} = e^{-3} = 0.367879441 - 0.049787068$

$= 0.318092372$

$= 0.318$

The length of the interval is $3X - X = 2X$

The expected length is $E(2X) = 2E(X)$

$E(X) = \int_0^\infty xe^{-x}dx$ set $u = x$ and $\frac{dv}{du} = e^{-x}$.

We have $\int \left(u\frac{dv}{du}\right)dx = uv - \int \left(u\frac{du}{dx}\right)dx$

$$= -xe^{-x} + \int_0^\infty e^{-x}dx$$

$= -xe^{-x} - e^{-x}$

$= [-e^{-x}(x+1)]_0^\infty = [e^{-x}(x+1)]_{\infty\infty}^0$

$= 1 - 0 = 1$

$\therefore \quad 2E(X) = 2 \times 1 = 2$

$\therefore$ Expected length $= 2$.

2. Confidence Interval for Means

Consider a sample $X_1, \dots, X_n$ from a distribution which is $N(\mu, \sigma^2)$, σ^2 known.

$$Z = (\bar{X} - \mu)/\sigma \tag{2.1}$$

is a unit normal variable, whatever the true value of μ may be. Hence we infer that

$$P(|Z|) \leq 1.96 = 0.954 \tag{2.2}$$

So that
$$P\left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + 1.96\frac{\sigma}{\sqrt{n}}\right) \tag{2.2.1}$$

Again,
$$T = \frac{(\bar{X} - \mu)\sqrt{n-1}}{s} \tag{2.3}$$

has a t distribution with (n – 1) degrees of freedom, whatever the value of $\sigma^2 > 0$, and $s^2 = \frac{1}{n}\sum_{1}^{n}(x_i - \bar{x})^2$. For a probability of 0.95, we can ind numbers $a < b$ from the table of t distribution such that

$$P\left(a < \frac{(\bar{X} - \mu)\sqrt{n-1}}{s} < b\right) = 0.95 \tag{2.4}$$

Noting that the random variable T is symmetric about the vertical axis through the origin, we would take $a = -b$ with $b > 0$. We have

$$P\left(\bar{X} - \frac{bs}{\sqrt{n-1}} < \mu < \bar{X} + \frac{bs}{\sqrt{n-1}}\right) = 0.95 \tag{2.5}$$

Example 3:

Let $n = 10$, $\bar{x} = 3.22$ and $s^2 = 1.3689$. Compute a 95% C.I. for μ.

Solution: From equation (2.5),

$b = t_{\alpha/2,9} = 2.262$ and $s = 1.17$.

We have $\left[3.22 - \frac{(2.262)(1.17)}{\sqrt{9}},\ 3.22 + \frac{(2.262)(1.17)}{\sqrt{9}}\right]$

or $(2.34,\ 4.10)$.

Example 4:

Let a random sample of size 20 from a distribution which is $N(\mu, 80)$ have mean $\bar{x} = 81.2$. Find a 95% confidence interval for μ.

Solution: $n = 20$, and $\sigma^2 = 80$

$$X \sim N(\mu, 80) \text{ and } \bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

The required interval is given by equation (2.2.1)

i.e., $\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}} = \bar{x} \pm 1.96(2)$

i.e. 81.2 ± 3.92

$= (77.28,\ 85.12)$

LECTURE SIX

UNIFORMLY MINIMUM VARIANCE UNBIASED ESTIMATOR (UMVUE)

CRAMER-RAO INEQUALITY

Consider the problem of how to find the best unbiased estimator of the parameter θ in the continuous density function $f(x;\theta)$. The solution of the problem lies in obtaining an inequality for the variance of any unbiased estimator $T = t(X_1, X_2, \dots, X_n)$ of θ. This inequality is derived in the following manner. Since $X_1, X_2, \dots, X_n$ is a random sample from $f(x;\theta)$, its density function will be denoted by L, where

$$L = \prod_{i=1}^{n} f(x;\theta)$$

It follows that

$$\int \dots \int L dx_1 dx_2 \dots dx_n = 1 \qquad (1.1)$$

Since $T = t(X_1, X_2, \dots, X_n)$ is assumed to be an unbiased estimator of θ, it follows that

$$E(t) = \int \dots \int +L dx_1 dx_2 \dots dx_n = \theta \qquad (1.2)$$

We differentiate (1.1) and (1.2) and assume that it is permissible to differentiate under the integral sign and that the limits of integration do not depend on θ. Differentiation of (1.1) will give

$$\int \dots \int \frac{\partial L}{\partial \theta} dx_1 dx_2 \dots dx_n = 0 \qquad (1.3)$$

Differentiation of (1.2) yields

$$\int \dots \int t \frac{\partial L}{\partial \theta} dx_1 dx_2 \dots dx_n = 1 \qquad (1.4)$$

The value of $\frac{\partial L}{\partial \theta}$ is most easily obtained by calculating

$$\frac{\delta \log L}{\delta \theta} = \frac{\delta \log L}{\delta L} \cdot \frac{\delta L}{\delta \theta} = \frac{1}{L} \frac{\delta L}{\delta \theta}.$$

Thus

$$\frac{\delta L}{\delta \theta} = L \sum_{i=1}^{h} \frac{\delta \log f(x_i;\theta)}{\delta \theta}$$

Let

$$T = \sum_{i=1}^{h} \frac{\delta \log f(x_i;\theta)}{\delta \theta} \tag{1.5}$$

Equation (1.2) can be expressed as

$$0 = \int \dots \int TL dx_1 dx_2 \dots dx_n = E(T) \tag{1.6}$$

Similarly, equation (1.4) will assume the form

$$1 = \int \dots \int tTL dx_1 dx_2 \dots dx_n = E[tT] \tag{1.7}$$

Next, consider the value of the correlation coefficient between the two random variables t and T. That is,

$$P_{tT} = \frac{E(tT)E(t)E(T)}{\sigma_t \sigma_T}$$

In view of the results in (1.6) and (1.7) this will reduce to

$$P_{tT} = \frac{1}{\sigma_t \sigma_T} \tag{1.8}$$

Since any correlation coefficient satisfies the inequality $P^2 \leq 1$, it follows from (1.8) that σ_t and σ_T must satisfy the inequality

$$\sigma_t^2 \geq \frac{1}{\sigma_T^2} \tag{1.9}$$

In view of (1.5) and the independence of the terms in that sum, it follows that

$$\sigma_T^2 = \sum_{i=1}^{n} \sigma_i^2 \tag{1.10}$$

where σ_i^2 is the variance of $\frac{\delta}{\delta \theta} f(x_i;\theta)$. But from (1.5) and (1.6)

$$\sum_{i=1}^{h} E \frac{\delta \log f(x_i;\theta)}{\delta \theta} = 0$$

Since the X_i possess the same distribution, the quantities $\frac{\delta \log f(x_i;\theta)}{\delta\theta}$, $i = 1,2,\dots,n$, must possess the same distribution, hence the same expected value. Since the sum of such expected values is zero, it follows that each expected value must be zero and therefore the variance σ_i^2 of $\frac{\delta \log f(x_i;\theta)}{\delta\theta}$ is equal to its second moment. Hence

$$\sigma_i^2 = E\left[\frac{\delta \log f(x_i;\theta)}{\delta\theta}\right]^2$$

Consequently from (1.10),

$$\sigma_T^2 = nE\left[\frac{\delta \log f(x_i;\theta)}{\delta\theta}\right]^2 \qquad (1.11)$$

because each X_i has the same distribution as the basic variable X. Substituting from (1.11) in (1.9) we have

$$\sigma_t^2 \geq nE\frac{1}{\left[\frac{\delta \log f(x_i;\theta)}{\delta\theta}\right]} \qquad (1.12)$$

Example: Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables from a Poisson distribution with parameter θ. Show that $\bar{X}$ is UMVU estimator of θ.

We have

$f(x;\theta) = e^{-\theta}\theta^x/x!$ so that

$log\ f(x;\theta) = -\theta + x\log\theta - \log x$

$\frac{\delta}{\delta\theta}\log f(x_i;\theta) = -1 + \frac{x}{\theta}$

$$E\left[\frac{\delta}{\delta\theta}\log f(x_i;\theta)\right]^2 = \left[1 - \frac{2x}{\theta} + \frac{x^2}{\theta^2}\right] = \frac{1}{\theta}$$

Since $E(x) = \theta$ and $E(x^2)\theta(1+\theta)$

The C-R lower bound = θ/A. Since $\bar{X}$ is unbiased for θ, with variance θ/n, we have that $\bar{X}$ is UMVE estimator of θ.

Exercises

1. Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables from a Bernoulli distribution, $B(1,\theta)$. Show that $\bar{x}$ is UMVU estimator of θ.

2. Let $X_1, X_2, \dots, X_n$ be i.i.d. random variables from $N(\mu, \sigma^2)$. Assume σ^2 is known and $\mu = \theta$. Show that $\bar{x}$ is a UMVU estimator of θ.

LECTURE SEVEN

TESTING HYPOTHESES

Statistical hypothesis is an assertion about the density function of a random variable. For example, given a probability density $f(x/\theta)$ and a sample $X_1, X_2, \ldots, X_n$ from it, a typical problem of testing a hypothesis, that is, a problem for which there are only two possible actions available, is to decide by means of a decision function $d = d(X_1, X_2, \ldots, X_n)$ whether $\theta \leq \theta_0$ or $\theta > \theta_0$, where θ_0 is some specified value. For example, the statement that the mean of a Poisson random variable is 5 is a statistical hypothesis.

Let us consider how a statistician proceeds in attempting to design a test that possesses desirable properties. Assume an exponential density is given by

$$f(x/\theta) = \theta e^{-\theta x} \qquad (1.1)$$

Assume further that the parameter θ has the value 2. This assumption is the statistical hypothesis to be tested, denoted by H_0. Let H_1 denote the alternative hypothesis that $\theta = 1$. Since there are only two possible actions that can be taken in this testing problem, namely accept H_0 or accept H_1, a decision function $d = d(X_1, X_2, \ldots, X_n)$ must separate n dimensional sample space into two parts. Let A_0 denote the part that is associated with accepting H_0, and A_1 the remaining part associated with accepting H_1. This means that if a random sample of X yields a point x=$(x_1, x_2, \ldots, x_n)$ that lies in A_0, we accept the hypothesis $H_0{:}\,\theta = \theta_0$, whereas if it lies in A_1, we accept the alternative hypothesis $H_1{:}\,\theta = \theta_1$. To avoid complicating the discussion at this stage, only one observation is taken on X. The problem of constructing a test for H_0 under discussion is therefore the problem of choosing a critical region on the positive x axis. This will lead to two types of error.

Two Types of Error

Suppose the statistician selects the part of the X axis to the right of $x = 1$ as the critical region. To decide whether this was a wise choice, we consider its consequences. If H_0 is actually true and the observed value of X exceeds 1, H_0 will be rejected. This, of course, is an incorrect decision. This type of error is called the type I error. On the other hand, if H_1 is actually true and the observed value of X does not exceed 1, H_0 will be accepted. This also is an incorrect decision. This kind of error is called the type II error. These two incorrect decisions, as well as the two correct decisions that are possible here, are displayed in table 1 below.

Table 1: Showing Two Types of Error

Status of H_0 and H_1 / Value of x	H_0 is True	H_1 is True
$x > 1$ (reject H_0)	Type I Error	Correct Decision
$x \leq 1$ (accept H_0)	Correct Decision	Type II Error

It is necessary to measure the seriousness of making either one of these errors before one can judge whether the choice of a critical region was wise. The size of an error is the measure of its seriousness. In the sequel, the loss function for our testing problem is given by

$$L(\theta, d) = \begin{cases} 0, & \text{if the correct decision is made} \\ 1, & \text{if an incorrect decision is made} \end{cases}$$

Definition 1:

α = size of type I error = P(Accept H_1/H_0 is true) i.e. the sample falls in the critical region, when in fact H_0 is true.

β = size of type II error = P(Accept H_0/H_1 is true)

Note that the two possible values of the risk function are given by

$$R(\theta_0, d) = EL\big(\theta_0, d(X)\big) = 0.P(X \in A_0/A_0) + 1.P(X \in A_1/\theta_0)$$

$$= P(X \in A_1/\theta_0) \qquad (1.2)$$

and

$$R(\theta_1, d) = EL\big(\theta_1, d(X)\big) = 1.P(X \in A_0/\theta_1) + 0.P(X \in A_1/\theta_1)$$

$$= P(X \in A_0/\theta_1) \qquad (1.3)$$

Definition 2:

The critical region of a test is that part of sample space that corresponds to the rejection of the hypothesis H_0. The size of the critical region, α, is the probability of the sample point falling in the critical region when H_0 is true.

Definition 3:

A best critical region of size α is one that minimizes the probability, β, of accepting H_0 when H_1 is true among all critical regions whose size does not exceed α. A best test is a test that is based on a best critical region.

Example: If X has the density $f(x/\theta) = \theta e^{-\theta x}$, $x > 0$, $\theta > 0$, and zero otherwise, if you are testing the hypothesis H_0: $\theta = 2$ against H_1: $\theta = 1$ by means of a single observed value of X and the critical region is $X \geq 1$, compute the sizes of α and β.

Solution:

$$\alpha = P(X \in \text{critical region}/\text{H}_0 \text{ is true})$$

$$= \int_1^{\infty} 2e^{-2x} dx = [e^{-2x}]_{\infty}^{1} = e^{-2} = 0.135335$$

$$\beta = P(X \in \text{feasible}/\text{H}_1 \text{ is true})$$

$= \int_0^1 e^{-x} dx = [e^{-1}]_1^0 = (1 - e^{-1})$

$= 1 - 0.36787944$

$= 0.632120$

Exercise:

Given $f(x;\theta) = \frac{1}{\theta}$, $0 \leq x \leq \theta$ and given the hypothesis $H_0\colon \theta = 1$ against the alternative $H_1\colon \theta = 2$, suppose a single observed value of *X* is to be taken.

(a) If the critical region is to be chosen to be the interval $X > \frac{1}{2}$, what is the values of α and β?

(b) What would those values become if $X > 1.5$ were chosen as the critical region?

(c) Comment on the power of the test in (a) above.