Unit I

Consistency

An estimator T based on the random observations of size n is said to be consistent estimator if it converges in probability to the true value of the parameter θ as $n \to \infty$.

More precisely, an estimator T is a consistent estimator of θ if

$$\lim_{n \to \infty} P[|T-\theta| > \epsilon] = 0$$

or equivalently

$$\lim_{n \to \infty} P[|T-\theta| < \epsilon] = 1 \quad \text{for } \epsilon > 0$$

Examples.

1. Show that the sample mean of normal distribution is a consistent estimator of μ, the population mean.

Sol: Let $x_1, x_2, x_3 \ldots x_n$ be a random sample from a normal population with mean μ and variance σ^2. i.e, $X \sim N(\mu, \sigma^2)$.

Let $\bar{x}$ be the sample mean. We know that $\bar{x} \sim N(\mu, \sigma^2/n)$.

Now

$$\lim_{n \to \infty} P[|\bar{x}-\mu| < \epsilon] = \lim_{n \to \infty} P[\mu - \epsilon < \bar{x} < \mu + \epsilon]$$

$$= \lim_{n \to \infty} \int_{\mu-\epsilon}^{\mu+\epsilon} f(\bar{x})\, d\bar{x}$$

$$= \lim_{n \to \infty} \frac{1}{\sigma/\sqrt{n} \cdot \sqrt{2\pi}} \int_{\mu-\epsilon}^{\mu+\epsilon} e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2} d\bar{x}$$

Let $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \Rightarrow \bar{x} - \mu = \frac{\sigma}{\sqrt{n}} z$

$$d\bar{x} = \frac{\sigma}{\sqrt{n}} \cdot dz$$

Lt: $\mu - \epsilon$ to $-\epsilon \cdot \frac{\sqrt{n}}{\sigma}$

$\mu + \epsilon$ to $\epsilon \cdot \frac{\sqrt{n}}{\sigma}$.

Substituting the above, we get

$$\lim_{n \to \infty} P[|\bar{x}-\mu| < \epsilon] = \lim_{n \to \infty} \frac{1}{\frac{\sigma}{\sqrt{n}} \cdot \sqrt{2\pi}} \int_{-\epsilon\frac{\sqrt{n}}{\sigma}}^{\epsilon \cdot \frac{\sqrt{n}}{\sigma}} e^{-\frac{z^2}{2}} \cdot \frac{\sigma}{\sqrt{n}} \cdot dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{z^2}{2}} \, dz$$

$\therefore \underset{n\to\infty}{Lt} \, P[|\bar{x} - \mu| < \epsilon] = 1.$

Hence $\bar{x}$ is a consistent estimator of μ.

2. Show that the sample median of a normal distribution is a consistent estimator of the population mean μ.

Sol: Let $x_1, x_2, \ldots x_n$ be a random sample drawn from a normal population with mean μ and variance σ^2. i.e. $X \sim N(\mu, \sigma^2)$.

Let $\tilde{x}$ be the sample median.

We know that $\tilde{x} \sim N(\mu, \frac{\sigma^2}{2n})$.

To prove, $\tilde{x}$ is a consistent estimator of μ, consider

$$\underset{n\to\infty}{Lt} \, P[|\tilde{x} - \mu| < \epsilon] = \underset{n\to\infty}{Lt} \, P[\mu - \epsilon < \tilde{x} < \mu + \epsilon]$$

$$= \underset{n\to\infty}{Lt} \int_{\mu-\epsilon}^{\mu+\epsilon} f(\tilde{x}) \, d\tilde{x}$$

$$= \underset{n\to\infty}{Lt} \frac{1}{\sigma\sqrt{\frac{\pi}{2n}}\sqrt{2\pi}} \int_{\mu-\epsilon}^{\mu+\epsilon} f(\tilde{x}) \, d\tilde{x}$$

Let $z = \frac{\tilde{x} - \mu}{\sigma \cdot \sqrt{\frac{\pi}{2n}}}$. $\Rightarrow \tilde{x} - \mu = \sigma \cdot \sqrt{\frac{\pi}{2n}} \, z$

$$\tilde{x} = \mu + \sigma \cdot \sqrt{\frac{\pi}{2n}} \cdot z$$

$$d\tilde{x} = \sigma \cdot \sqrt{\frac{\pi}{2n}} \cdot dz$$

Lt: $\mu - \epsilon \Rightarrow -\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}$

$\mu + \epsilon \Rightarrow \frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}$

$$\therefore \underset{n\to\infty}{Lt} \, P[|\tilde{x} - \mu| < \epsilon] = \underset{n\to\infty}{Lt} \frac{1}{\sigma\sqrt{\frac{\pi}{2n}} \cdot \sqrt{2\pi}} \int_{-\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}}^{\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}} e^{-\frac{z^2}{2}} \cdot \sigma\sqrt{\frac{\pi}{2n}} \cdot dz$$

$$= \underset{n\to\infty}{Lt} \frac{1}{\sqrt{2\pi}} \int_{-\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}}^{\frac{\epsilon\sqrt{2n}}{\sigma\sqrt{\pi}}} e^{-z^2/2} \, dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} \, dz$$

$$= 1$$

$\therefore \tilde{x}$ is a consistent estimator of μ.

Scanned with CamScanner

3

$x_1, x_2 \ldots x_n$ be a random sample drawn from a Cauchy population with parameter θ. Show that the sample mean is an inconsistent estimator of θ.

Sol: The cauchy distribution with parameter θ is given by

$$f(x) = \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}, \quad -\infty < x < \infty.$$

To verify, whether $\bar{x}$ is a consistent estimator of θ, consider

$$\lim_{n\to\infty} P[|\bar{x}-\theta|<\epsilon] = \lim_{n\to\infty} \int_{\theta-\epsilon}^{\theta+\epsilon} f(\bar{x})\, d\bar{x}$$

$$= \lim_{n\to\infty} \frac{1}{\pi} \int_{\theta-\epsilon}^{\theta+\epsilon} \frac{1}{1+(\bar{x}-\theta)^2}\, d\bar{x}$$

Let $z = \bar{x}-\theta$, we get
$dz = d\bar{x}$

Lt: $\theta-\epsilon \Rightarrow -\epsilon$ $\quad \theta+\epsilon \Rightarrow \epsilon$

$$\therefore \lim_{n\to\infty} P[|\bar{x}-\theta|<\epsilon] = \lim_{n\to\infty} \frac{1}{\pi} \int_{-\epsilon}^{\epsilon} \frac{1}{1+z^2}\, dz$$

$$= \lim_{n\to\infty} \frac{2}{\pi} \int_{0}^{\epsilon} \frac{1}{1+z^2}\, dz$$

$$= \lim_{n\to\infty} \frac{2}{\pi} \cdot [\tan^{-1} z]_0^{\epsilon}$$

$$= \lim_{n\to\infty} \frac{2}{\pi} [\tan^{-1}\epsilon - 0] \neq 1$$

Since $\lim_{n\to\infty} P[|\bar{x}-\theta|<\epsilon] \neq 1$, we can say that the sample mean of Cauchy distribution is an inconsistent estimator of θ.

4. Show that the sample median of Cauchy distribution is a consistent estimator of θ.

Sol: Let $x_1, x_2 \ldots x_n$ be a random sample drawn from a cauchy population. Let $\tilde{x}$ be the sample median. We know that $\tilde{x} \sim N(\theta, \frac{\pi^2}{4n})$.

To prove $\tilde{x}$ is a consistent estimator of θ, consider

$$\lim_{n\to\infty} P[|\tilde{x}-\theta|<\epsilon] = \lim_{n\to\infty} \int_{\theta-\epsilon}^{\theta+\epsilon} f(\tilde{x})\, d\tilde{x}.$$

$$= \lim_{n \to \infty} \frac{1}{\frac{\pi}{\sqrt{4n}} \sqrt{2\pi}} \int_{\theta-\epsilon}^{\theta+\epsilon} e^{-\frac{1}{2}\left(\frac{\tilde{x}-\theta}{\pi/\sqrt{4n}}\right)^2} \cdot d\tilde{x}$$

Let $z = \frac{\tilde{x} - \theta}{\pi/\sqrt{4n}}$

$\Rightarrow dz = \frac{d\tilde{x}}{\pi/\sqrt{4n}} \Rightarrow d\tilde{x} = \frac{\pi}{\sqrt{4n}} \cdot dz$

Lt : $\theta - \epsilon$ to $\theta + \epsilon \Rightarrow -\frac{\epsilon\sqrt{4n}}{\pi} \to \frac{\epsilon\sqrt{4n}}{\pi}$

$$\therefore \lim_{n \to \infty} P\left[\,|\tilde{x} - \theta| < \epsilon\,\right] = \lim_{n\to\infty} \frac{1}{\frac{\pi}{\sqrt{4n}}\sqrt{2\pi}} \int_{-\frac{\epsilon\sqrt{4n}}{\pi}}^{\frac{\epsilon\sqrt{4n}}{\pi}} e^{-\frac{z^2}{2}} \cdot \frac{\pi}{\sqrt{4n}} \cdot dz$$

$$= \lim_{n\to\infty} \frac{1}{\sqrt{2\pi}} \int_{-\frac{\epsilon\sqrt{4n}}{\pi}}^{\frac{\epsilon\sqrt{4n}}{\pi}} e^{-\frac{z^2}{2}} dz$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-z^2/2} dz = 1$$

$\therefore$ Sample median is a consistent estimator of θ.

Sufficient condition for consistency:

The statistic T is said to be consistent estimator for the parameter θ if

i) $E(T) = \theta_n \to \theta$
ii) $V(T) \to 0$

} as $n \to \infty$

Remark:

Using the above sufficient condition for consistency, we can easily prove whether an estimator is consistent or not. But Cauchy distribution is an exceptional case. The reason is that the sample mean of a Cauchy distribution follows Cauchy distribution with parameter θ and we know that the moments of Cauchy distribution do not exist. Hence this condition may not be used to examine whether the sample mean of Cauchy distribution is consistent or not.

However, a statistic, sample median of Cauchy distribution for large n is distributed like normal distribution with mean θ and variance $\frac{\pi^2}{4n}$. Hence by applying the sufficient condition for consistency we can prove that the sample median is a consistent estimator of θ.

Example

1. Show that the sampling variance s^2 of normal population is a consistent estimator for the population variance σ^2.

Sol: We know that

$$E(\chi^2) = n-1 \quad \text{and} \quad V(\chi^2) = 2(n-1)$$

ie $$E\left(\frac{ns^2}{\sigma^2}\right) = n-1 \quad \text{and} \quad V\left(\frac{ns^2}{\sigma^2}\right) = 2(n-1)$$

$$\therefore E(s^2) = \frac{n-1}{n}.\sigma^2$$

$$V(s^2) = \frac{2(n-1)}{n^2}.\sigma^4$$

By the sufficient condition for consistency, we have as $n \to \infty$

$$E(s^2) = \lim_{n\to\infty}\left(\frac{n-1}{n}\right).\sigma^2 = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)\sigma^2 = \sigma^2$$

$$V(s^2) = \lim_{n\to\infty} \frac{2(n-1)}{n^2}.\sigma^4$$

$$= \lim_{n\to\infty} \frac{2\,n(1-\frac{1}{n})}{n \cdot n}\,\sigma^4$$

$$\to 0 \text{ as } n \to \infty.$$

Hence s^2 is a consistent estimator of σ^2.

2. Show that the sample mean $\bar{x}$ of a population with mean θ and variance σ^2 is always a consistent estimator of θ (irrespective of the population).

note: Cauchy distribution is exceptional one.

Sol: Let $x_1, x_2 \ldots x_n$ be a random sample selected from a normal population with mean θ and variance σ^2.

By def, we have $$E(\bar{x}) = E\left[\frac{x_1+x_2+\ldots+x_n}{n}\right]$$

$$= \frac{1}{n}\left[E(x_1) + E(x_2) + \dots + E(x_n)\right]$$

Since $E(x_i) = \theta \quad \forall\, i,\ i = 1, 2 \dots n$

$$E(\bar{x}) = \frac{1}{n}\left[\theta + \theta + \dots + \theta\right] = \frac{n\theta}{\theta} = \theta.$$

Considering the variance, we have

$$V(\bar{x}) = V\left(\frac{\Sigma x_i}{n}\right)$$

$$= \frac{1}{n^2} V(\Sigma x_i)$$

$$= \frac{V\left[x_1 + x_2 + \dots + x_n\right]}{n^2}$$

$$= (\sigma^2 + \sigma^2 + \dots + \sigma^2)/n^2 = \frac{\sigma^2}{n}$$

$$\lim_{n \to \infty} \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty.$$

Hence $\bar{x}$ is a consistent estimator of θ, using the sufficient condition for consistency.

2. Unbiasedness

The property of consistency is a limiting property. Consistency is defined for only infinite sample size and it becomes invalid for any specific size. Sometimes, we may have two or more consistent estimators for the same parameter.

For example, it is known that the sample mean as well as sample median are both consistent for the population mean in the case of normal distribution. Moreover, if T is a consistent estimator for θ, for any arbitrary constants, $a, b, c \dots$ the statistics

$$\frac{n+a}{n-b} T, \quad \frac{n^2 - a^2}{n^2 + a + b} T \dots$$ are all consistent estimators

for θ. Hence inorder to find an optimum estimator among the class of estimators, some other criterion is needed for which the concept of unbiasedness is introduced.

Definition: Any statistic whose mathematical expectation is equal to a parameter θ is called an unbiased estimator of the parameter θ. Otherwise, the statistic is said to be biased.

The statistic T will be called an unbiased estimator of $T\theta$, if it has 0 bias for all θ.

More precisely, the estimator T is said to be unbiased for the parameter θ if $E(T) = \theta$.

T is said to be asymptotically unbiased if

$$\lim_{n \to \infty} E(T) = \theta.$$

Remark:

i) The criteria for unbiasedness also does not satisfy uniqueness characteristics.

For example, in normal distribution, both sample mean and median are consistent estimator for μ.

ii) The criteria of consistency and unbiasedness neither imply each other.

For example, in Cauchy distribution, sample mean is not a consistent estimator.

Example:

1. If T is an unbiased estimator of θ, prove that T^2 is a biased estimator of θ^2.

Sol: It is given that $E(T) = \theta$. We have to prove that $E(T^2) \neq \theta^2$.

Consider $V(T)$.

By def $V(T) = E(T^2) - [E(T)]^2$

Since variance is always ≥ 0, we can write

$$E(T^2) - [E(T)]^2 \geq 0$$

$$\Rightarrow E(T^2) - \theta^2 \geq 0$$

$$\Rightarrow E(T^2) \geq \theta^2.$$

$\Rightarrow T^2$ is not an unbiased estimator of θ^2.

2. Let $x_1, x_2, \ldots x_n$ be a random sample drawn from a normal population with mean μ and variance 1. Show that $\frac{1}{n}\Sigma x_i^2$ is an unbiased estimator of $\mu^2 + 1$.

$x_1, x_2 \ldots x_n$ be a random sample drawn from a normal population with mean μ and variance 1. Show that $\frac{1}{n}\sum_{1}^{n} x_i^2$ is an unbiased estimator of $T_\mu = \mu^2+1$.

Sol: It is given that $X \sim N(\mu, 1)$

By def $V(x_i) = E(x_i^2) - [E(x_i)]^2$

$= E(x_i^2) - \mu^2$

ie $1 = E(x_i^2) - \mu^2$

$\Rightarrow E(x_i^2) = \mu^2 + 1$

Consider $E(\bar{x}) = E\left[\frac{1}{n}\sum x_i\right]$

$= \frac{1}{n}\sum E(x_i) = \frac{1}{n}\sum\theta = \theta.$

$\therefore \bar{x}$ is an unbiased estimator of θ.

4. Show that the sample variance $s^2 = \frac{1}{n-1}\sum(x-\bar{x})^2$ is an unbiased estimator of σ^2, where σ^2 is the population variance (known).

Sol:

$$E(s^2) = \frac{E1}{n-1}\left[\frac{1}{n-1}\sum x - \bar{x})^2\right]$$

$$= \frac{E}{n-1}\left[\frac{1}{n-1}\sum x - \mu + \mu - \bar{x})^2\right]$$

$$= \frac{E}{n-1}\left[\sum(x-\mu)^2 - n(\bar{x}-\mu)^2\right]$$

$$= \frac{\sum E(x-\mu)^2}{n-1} - \frac{n}{n-1}E(\bar{x}-\mu)^2$$

$$= \frac{\sum V(x_i)}{n-1} - \frac{n}{n-1}\cdot\frac{\sigma^2}{n}$$

$$= \frac{n}{n-1}\cdot\sigma^2 - \frac{\sigma^2}{n-1}$$

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

$\therefore s^2$ is an unbiased estimator of σ^2.

Note: The sample variance $s^2 = \frac{\sum(x-\bar{x})^2}{n-1}$ is an asymptotic unbiased estimator of σ^2 since

$$\lim_{n\to\infty} E(s^2) = \lim_{n\to\infty}\frac{n-1}{n}\cdot\sigma^2$$

$$= \sigma^2.$$

Invariance property of consistent estimator:

If T_n is a consistent estimator of $\gamma(\theta)$ and $\psi[\gamma(\theta)]$ is a continuous function of $\gamma(\theta)$, then $\psi(T_n)$ is a consistent estimator of $\psi[\gamma(\theta)]$

Proof: Since T_n is a consistent estimator of $\gamma(\theta)$,

$$T_n \rightarrow \gamma(\theta) \text{ as } n \rightarrow \infty$$

ie for every $\epsilon > 0$, $\eta > 0$, there exist a positive integer $n \geq m(\epsilon, \eta)$ such that

$$P\{|T_n - \gamma(\theta)| < \epsilon\} > 1 - \eta, \quad \forall\, n \geq m$$

Since $\psi(.)$ is a continuous function, for every $\epsilon > 0$, however small, there exist a positive number ϵ_1 such that

$$|\psi(T_n) - \psi[\gamma(\theta)]| < \epsilon_1 \text{ whenever } |T_n - \gamma(\theta)| < \epsilon$$

ie $|T_n - \gamma(\theta)| < \epsilon \Rightarrow |\psi(T_n) - \psi\,\gamma(\theta)| < \epsilon_1$ — ①

For two events A and B if $A \Rightarrow B$, then $P(B) \geq P(A)$ — ②

$A \subseteq B \Rightarrow P(A) \leq P(B)$ or

From ① and ② we get

$$P\{|\psi(T_n) - \psi(\gamma(\theta))| < \epsilon_1\} \geq P\{|T_n - \gamma(\theta)| < \epsilon\}$$

$$\Rightarrow P\{|\psi(T_n) - \psi\,\gamma(\theta)| < \epsilon_1\} \geq 1 - \eta, \quad \forall\, n \geq m.$$

$$\Rightarrow \psi(T_n) \xrightarrow{P} \psi[\gamma(\theta)] \text{ as } n \rightarrow \infty$$

or $\psi(T_n)$ is a consistent estimator of $\gamma(\theta)$.

Theorem. 2 Sufficient condition for consistency.

Let $\{T_n\}$ be a sequence of estimators such that for all $\theta \in \Theta$

i) $E_\theta(T_n) \rightarrow \gamma(\theta)$ as $n \rightarrow \infty$

ii) $V_\theta(T_n) \rightarrow 0$ as $n \rightarrow \infty$.

Proof: We have to prove that T_n is a consistent estimator of $\gamma(\theta)$.

That is $T_n \longrightarrow \gamma(\theta)$ as $n \rightarrow \infty$.

ie $P[|T_n - \gamma(\theta)| < \epsilon] > 1 - \eta \quad \forall\, n \geq m(\epsilon, \eta)$ — ①

where ϵ and η are arbitrary small positive numbers and m is some large value of n

...ying Chebyshev's inequality, to a statistic T_n, we get

$$P[\,|T_n - E_\theta(T_n)| \le \delta\,] \ge 1 - \frac{var(T_n)}{\delta^2} \quad \text{— (2)}$$

we have

$$|T_n - \gamma(\theta)| = |T_n - E(T_n) + E(T_n) - \gamma(\theta)| \quad \text{— (3)}$$

$$\le |T_n - E_\theta(T_n)| + |E_\theta(T_n) - \gamma(\theta)| \quad \text{— (4)}$$

[illegible]

$$|T_n - E_\theta(T_n)| \le \delta \Rightarrow$$

$$\Rightarrow |T_n - \gamma(\theta)| \le \delta + |E_\theta(T_n) - \gamma(\theta)| \quad \text{— (5)}$$

Hence using (2) of previous theorem, we get

$$P\{|T_n - \gamma(\theta)| \le \delta + |E_\theta(T_n) - \gamma(\theta)|\}$$

$$\ge P\{|T_n - E_\theta(T_n)| \le \delta\}$$

$$\ge 1 - \frac{var_\theta(T_n)}{\delta^2} \quad \text{from (2)}$$

we are given

$$E_\theta(T_n) \to \gamma(\theta) \quad \forall\ \theta \quad (4) \quad \text{as } n \to \infty$$

Hence for every $\delta_1 > 0$ $\exists$ a positive integer $n \ge n_0(\delta_1)$ such that

$$|E_\theta(T_n) - \gamma(\theta)| \le \delta_1 \quad \forall\ n \ge n_0(\delta_1) \quad \text{— (6)}$$

Also $var_\theta(T_n) \to 0$ as $n \to \infty$ (given)

$$\therefore \frac{var_\theta(T_n)}{\delta^2} \le \eta, \quad \forall\ n \ge n_0(\eta) \quad \text{— (7)}$$

where η is arbitrary and small positive number.

Substituting (6) & (7) in (5), we get

$$P\{|T_n - \gamma(\theta)| \le \delta + \delta_1] \ge 1 - \eta; \quad n \ge m(\delta_1, \eta)$$

$$\Rightarrow P\{|T_n - \gamma(\theta)| \le \epsilon\} \ge 1 - \eta, \quad n \ge m$$

$$\epsilon = \delta + \delta_1 > 0$$

where $m = \max(n_0, n_0')$ and

$$\Rightarrow T_n \xrightarrow{P} \gamma(\theta) \quad \text{as } n \to \infty$$

$\therefore$ T_n is a consistent estimate of $\gamma(\theta)$

...ciency:

An estimator T is said to be sufficient for the parameter θ, if the likelihood function $L(x_1, x_2 \ldots x_n; \theta)$ can be written in the form

$$L(x_1, x_2 \ldots x_n; \theta) = g(T, \theta) \cdot h(x_1, x_2 \ldots x_n / T)$$

where $g(T, \theta)$ is the marginal density of T and $h(x_1, x_2 \ldots x_n / T)$ is the conditional density of $x_1 . x_2 .. x_n / T$ and it does not include θ.

Factorization Theorem:

The necessary and sufficient condition for a distribution to admit sufficient statistic is provided by the factorization theorem due to Neyman.

Statement:

$T = t(x)$ is sufficient for θ if and only if the joint density function L (say) of the sample values can be expressed in the form:

$$L = g_{\theta}[t(x)] \cdot h(x)$$

where $g_{\theta}(t(x))$ depends on θ and x only through the value of $t(x)$ and $h(x)$ is independent of θ.

Illustration for Sufficient Statistic.

Let us consider the problem of estimating p where p denotes the probability of getting heads when the coins are tossed.

Consider the problem of tossing 3 coins. To estimate for p, we need only the no. of heads obtained in the tosses. We do not need the exact sequence in which H and T are obtained.

we denote $x_i = 1$ if the toss results in H
$= 0$ if " " " T

∴ The no. of heads in 3 tosses $= t = \Sigma x_i$

If $t = 2$, what is the joint distribution of $(x_1\ x_2\ x_3)$ or in other words, what is the conditional probability of $x_1 x_2 x_3 / t = 2$

obtain the conditional probability, we write

$$P(HHT / t=2) = \frac{P(HHT, t=2)}{P(t=2)}$$

$$= \frac{p^2(1-p)}{{}^3C_2\, p^2 (1-p)} = 1/3.$$

$\therefore$ the conditional probability of $x_1, x_2 \ldots x_3 / t=2$ is independent of p. In this case, we say that t is sufficient of p. In otherwords it contains all the information about p.

Now, take the general case of tossing a coin n times. To show the conditional probability of $x_1, x_2 \ldots x_n / t = t_0$ does not involve p, consider the following probabilities.

i) The probability of getting $t = t_0$ is

$$P(t = t_0) = {}^nC_{t_0} \cdot p^{t_0} \cdot q^{n-t_0}$$

ii) The joint probability of $x_1, x_2 \ldots x_n$ is

$$P[X_1 = x_1,\ X_2 = x_2 \ldots X_n = x_n ;\ T = t_0 (\Sigma x_i)]$$

$$= p^{x_1} (1-p)^{1-x_1}\, p^{x_2} (1-p)^{1-x_2} \ldots p^{x_n} (1-p)^{(1-x_n)}$$

$$= p^{\Sigma x_i} (1-p)^{n - \Sigma x_i} \quad \text{if } \Sigma x_i = t_0$$

$$= 0 \quad \text{if } \Sigma x_i = 0$$

Hence the conditional probability

$$P[X_1 = x_1 \ldots X_n = x_n / t = t_0] = \frac{p^{t_0} (1-p)^{n-t_0}}{{}^nC_{t_0}\, p^{t_0} (1-p)^{n-t_0}}$$

$$= \frac{1}{{}^nC_{t_0}}$$

Thus the conditional distribution of $x_1 x_2 \ldots x_n / T$ is independent of p. This implies that the likelihood function can be written in the form

$$L(x_1, x_2 \ldots x_n ; p) = g(t, p)\, h(x_1 \ldots x_n / t)$$

Hence t is sufficient for p.

Scanned with CamScanner

...ler's Neyman criterion.

A statistic $t_1 = t(x_1, x_2 \ldots x_n)$ is sufficient estimator of parameter θ if and only if the likelihood function (joint p.d.f of the sample) can be expressed as

$$L = \prod_{i=1}^{n} f(x_i, \theta) = g_1(t_1, \theta) \cdot k(x_1, x_2 \ldots x_n)$$

where $g(t_1, \theta)$ is the p.d.f of the statistic t_1 and $k(x_1, x_2 \ldots x_n)$ is a function of sample observations only, independent of θ.

Example

1. Let $x_1, x_2 \ldots x_n$ be a random sample from a discrete population with distribution function

$$f(x) = \theta^x (1-\theta)^{1-x} \quad \text{if } x = 1 \text{ or } 0$$
$$= 0 \quad \text{otherwise.}$$

Obtain sufficient statistic for θ.

Sol: Let $x_1, x_2 \ldots x_n$ be a random sample taken from a population with p.m.f

$$f(x;\theta) = \theta^x (1-\theta)^{1-x} c(x), \quad \text{if } x = 1 \text{ or } 0$$
$$= 0 \quad \text{otherwise}$$

where $c(x) = 1$ if $x = 1$ or 0
$= 0$ otherwise.

Hence the likelihood function can be written as

$$L(x_1, x_2 \ldots x_n; \theta) = \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i} \prod_{i=1}^{n} c(x_i) \quad \text{if } x = 1 \text{ or } 0$$
$$= 0 \quad \text{otherwise.}$$

Now L can be factorized as

$$L = \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i} \, \Pi \, \Sigma x_i \quad \text{if } x = 1 \text{ or } 0$$

where $g(t_1, \theta) = \theta^{\Sigma x_i} (1-\theta)^{n-\Sigma x_i}$

$h(x_1 \ldots x_n) = \Pi \, c(x_i)$

Since $h(x_1, x_2 \ldots x_n)$ is independent of θ, applying Neyman factorization theorem, we can say that $T = \Sigma x_i$ is sufficient for θ.

Moreover, any one-one function of T is also sufficient.

$\Rightarrow \frac{T}{n} = \frac{\Sigma x_i}{n}$: sample proportion is sufficient for θ.

Scanned with CamScanner

. $x_1, x_2 \ldots x_n$ be a random sample drawn from normal population with mean μ and variance 1.

Sol: Let $x_1, x_2 \ldots x_n$ be a random sample drawn from a normal population with mean μ and variance σ^2.

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

The likelihood function is given by

$$L(x_1, x_2 \ldots x_n; \mu) = \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{1}{2}\Sigma(x - \mu)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\Sigma x^2}{2} - \frac{n\mu^2}{2}} + e^{\Sigma x \mu}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(e^{-n\mu^2 + \mu \Sigma x}\right) \cdot \left(e^{\frac{\Sigma x^2}{2}}\right)$$

$$= g(T, \theta) \cdot h(x)$$

Hence Σx is sufficient for μ.

Assignment

1. If $x_1, x_2 \ldots x_n$ is a random sample drawn from Poisson population with parameter λ. Show that $\bar{x}$ is the sufficient estimator for λ.

2. Let $x_1, x_2 \ldots x_n$ be a random sample drawn from exponential distribution with p.d.f

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

Obtain sufficient statistic for θ.

3. If $x_1, x_2 \ldots x_n$ is a random sample drawn from Cauchy population with parameter θ, examine if a sufficient statistic for θ. If so, obtain its value

Sufficiency for several variables.

Def: Let there be k parameters $\theta_1, \theta_2 \ldots \theta_k$ to be estimated. Let $T_1, T_2 \ldots T_k$ be k statistics. The statistic $T_1, T_2 \ldots T_k$ are said to be jointly sufficient for $\theta_1, \theta_2 \ldots \theta_k$ if the likelihood function can be written in the form

$$L(x_1, x_2 \ldots x_n; \theta_1, \theta_2 \ldots \theta_k) = g(T_1, T_2 \ldots T_k; \theta_1, \theta_2 \ldots \theta_k) \cdot h(x_1, x_2 \ldots x_n).$$

Example.

Obtain the sufficient estimator for μ and σ^2.

Sol:

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

Let $\mu = \theta_1$ and $\sigma^2 = \theta_2$

$$\therefore f(x) = \frac{1}{\sqrt{2\pi}\sqrt{\theta_2}} e^{-\frac{1}{2\theta_2}(x-\theta_1)^2}$$

$$L(x_1, x_2 \ldots x_n; \theta_1, \theta_2^2) = \left(\frac{1}{\sqrt{2\pi}\sqrt{\theta_2}}\right)^n e^{-\frac{1}{2\theta_2}\sum_{i=1}^{n}(x_i - \theta_1)^2}$$

$$= \left(\frac{1}{2\pi\theta_2}\right)^{n/2} e^{-\frac{1}{2\theta_2}\left[\sum(x-\bar{x})^2 + n(\bar{x}-\theta_1)^2\right]}$$

$$= g(\bar{x}, s^2; \theta_1, \theta_2)\, h(x_1 \ldots x_n)$$

where $s^2 = \frac{1}{n}\sum(x-\bar{x})^2$ and $h(x_1 . x_2 \ldots x_n) = 1$.

Hence $\bar{x}$ and s^2 are jointly sufficient for $(\theta_1\ \theta_2)$.

Assignment

If x_i $(i = 1, 2 \ldots n)$ is a random sample drawn from Beta population with parameters θ_1 and θ_2.

i) Obtain sufficient statistic for θ_1 when θ_2 is unknown.

ii) Obtain sufficient statistic for θ_2 when θ_1 is unknown.

Example.

1. Let $x_1, x_2 \ldots x_n$ be a random sample drawn from a uniform population ~~with~~ on $[0, \theta]$. Find sufficient estimator for θ.

Sol: Given

$$f_\theta(x_i) = \frac{1}{\theta}, \quad 0 \le x_i \le \theta$$
$$= 0 \quad \text{otherwise.}$$

$$L = \prod_{i=1}^{n} f(x_i, \theta) = \frac{1}{\theta^n}, \quad 0 < x_i < \theta$$

If $t = \max(x_1 \ldots x_n) = x_n$

then p.d.f of t is given by

$$g(t, \theta) = n\left[F(x_n)\right]^{n-1} f(x_{(n)})$$

We have $F(x) = P[X \le x] = \int_0^x f(x, \theta)\, dx$

$$= \int_0^x \frac{1}{\theta} \cdot dx = \frac{x}{\theta}.$$

$$g(t,\theta) = n\left[\frac{x_{(n)}}{\theta}\right]^{n-1} \cdot \frac{1}{\theta} = \frac{n}{\theta^n}\left[x_{(n)}\right]^{n-1}$$ 15

$$\therefore L = \frac{n\left[x_{(n)}\right]^{n-1}}{\theta^n} \cdot \frac{1}{n\left[x_{(n)}\right]^{n-1}} = g(t,\theta) \cdot h(x_1 \dots x_n)$$

Hence by Fisher - Neymann Criterion, the statistic $t = x_{(n)}$ is sufficient estimator of θ.

Assignment :

1. Let $x_1, x_2 \dots x_n$ be a random sample from a distribution with p.d.f

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

Obtain sufficient statistic for θ.

2. Let $x_1, x_2 \dots x_n$ be a random sample from a population with p.d.f

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

Show that $t_1 = \prod_{i=1}^{n} x_i$ is sufficient for θ.

Minimal sufficient statistic.

When we introduced the concept of sufficiency, our object was to condense the data without losing any information about the parameter. We have seen that there is more than one set of sufficient statistics.

For example, in sampling from a normal distribution with both mean and variance unknown, we have noted three sets of jointly sufficient statistics, namely, the sample $X_1, X_2 \ldots X_n$ itself, the order statistics $Y_1, Y_2 \ldots Y_n$ and $\bar{x}$ and S^2.

We naturally prefer the jointly sufficient set $\bar{x}$ and S^2 since they condense the data more than either of the two. Now the question is Does there exists a set of sufficient statistics that condenses the data more than $\bar{x}$ and S^2? The answer is that there does not, but we will not develop the necessary tools to establish the answer. The idea is the concept of minimum set of sufficient statistics which we call as minimal sufficient statistics.

We know that corresponding to any statistic is the partition of $\mathcal{X}$ that it induces. The same is true of a set of statistics: a set of statistics induces a partition of $\mathcal{X}$. The condensation of the data that a statistic or a set of statistics exhibits can measured by the number of subsets in the partition induced by that statistic or set of statistics.

If a set of statistics has fewer subsets in its induced partition than does the induced partition of another set of statistics, then we say that the first statistics condenses the data more than the latter. That is a minimal sufficient set of statistics is a sufficient set of statistics than the that has fewer subsets in its partition than the induced partition of any other set of sufficient statistics. So a set of sufficient statistics is minimal if no other set of sufficient statistics condenses the data more.

Minimal sufficient statistic.

A set of jointly sufficient statistics is defined to be minimal sufficient if and only if it is a function of every other set of sufficient statistic.

Distributions Admitting sufficient statistic.

Let the distribution have pdf or pmf $f(x, \theta)$ depending on a single parameter θ such that Θ is a real interval.

If there exist a sufficient statistic t for θ, according to factorization theorem, the likelihood function can be written as

$$L = g(t, \theta) \cdot h(x).$$

that is

$$\prod f(x) = g(t, \theta)\, h(x).$$

Taking log on both sides

$$\Sigma \log f(x) = \log g(t, \theta) + \log h(x).$$

Assuming that the functions f and g are partially differentiable w.r.t. θ, we get

$$\frac{\Sigma\, \partial \log f(x, \theta)}{\partial \theta} = \frac{\partial \log g(t, \theta)}{\partial \theta} = k(t, \theta) \text{ (say)} \quad \text{— (1)}$$

Since the equation holds for all θ, for any particular value of θ, the value of zero may be substituted in (1), we have

$$v = k(t) \quad \text{— (2)}$$

where $v = \Sigma\, v(x_i)$

$$v(x_i) = \left.\frac{\partial \log f(x_i, \theta)}{\partial \theta}\right|_{\theta = 0} \quad \text{and} \quad k(t, \theta)\big|_{\theta = 0} = k(t).$$

Suppose, $k(t)$ and v are differentiable, functions, then we have

$$\frac{\partial v}{\partial x_i} = \frac{\partial v(x_i)}{\partial x_i} = \frac{\partial k(t)}{\partial t} \cdot \frac{\partial t}{\partial x_i} \quad \text{— (3)}$$

From (1), on differentiating w.r.t. x_i, we get

$$\frac{\partial^2 \log f(x_i, \theta)}{\partial \theta\, \partial x_i} = \frac{\partial k(t, \theta)}{\partial t} \cdot \frac{\partial t}{\partial x_i} \quad \text{— (4)}$$

$$\frac{\partial^2 \log f(x_i, \theta)}{\partial\theta \cdot \partial x_i} = \frac{\partial k(t,\theta)}{\partial t} \cdot \frac{\partial t}{\partial x_i} \quad \text{— (4)}$$

Hence $$\frac{\partial^2 \log f(x_i; \theta)}{\partial\theta \cdot \partial x_i} \Big/ \frac{\partial v(x_i)}{\partial x_i} = \frac{\partial k(t, \theta)}{\partial t} \Big/ \frac{\partial k(t)}{\partial t}.$$

However, the LHS being the same $\forall x_i$, must depend on θ alone, so that for the RHS also, we have

$$\frac{\partial k(t,\theta)}{\partial k(t)} = \lambda(\theta) \text{ (say)}$$

Integrating we get

$$k(t,\theta) = \lambda_1(\theta) \cdot k(t) + \lambda_2(\theta).$$

But $$k(t,\theta) = \Sigma \frac{\log f(x_i; \theta)}{\partial\theta}$$

$$\therefore \Sigma \frac{\partial \log f(x_i; \theta)}{\partial\theta} = \lambda_1(\theta) \cdot k(t) + \lambda_2(\theta)$$

Integrating, we have

$$\Sigma \log f(x_i, \theta) = k(t) \int \lambda_1(\theta)\, d\theta + \int \lambda_2(\theta)\, d\theta + c_1(x_1 \cdots x_n)$$

$$\Sigma \log f(x, \theta) = A(\theta)\, \Sigma v(x) + B_1(\theta) + \Sigma \Sigma(x_i)$$

$$\Rightarrow f(x, \theta) = \exp\left[A(\theta)\, v(x) + B(\theta) + c(x)\right]$$

This is precisely the form of the exponential family of distribution to which most of the common distribution belong. eg. Binomal, Poisson, normal etc.

Hence the sufficient statistic for θ is $\Sigma v(x_i)$ or any one-one function of it.

...nential Family

A one-parameter family (θ) of densities ($f \cdot ; \theta$) that can be expressed as

$$f(\cdot;\theta) = a(\theta)\, b(x) \exp\left[c(\theta)\, d(x)\right], \quad \text{for } -\infty < x < \infty, \; \theta \in \Theta \quad \rightarrow (1)$$

and for a suitable choice of functions $a(\cdot), b(\cdot), c(\cdot)$ and $d(\cdot)$ is defined to belong to the exponential family or exponential class.

Example. 1 verify whether the following density belongs to the exponential family.

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

Sol: consider the density function

$$f(x;\theta) = \theta e^{-\theta x}$$

$$= \theta \cdot I_{(0,\infty)}(x)\, e^{-\theta \cdot x}$$

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

where $a(\theta) = \theta$, $b(x) = I_{(0,\infty)}(x)$

Example 2 verify whether Poisson distribution belongs to the exponential family.

Sol: consider

$$p(x) = \frac{e^{-\lambda} \lambda^{x}}{x!}$$

$$= \frac{e^{-\lambda} \lambda^{x}}{x!} I_{(0,1,\ldots)}(x)$$

$$= e^{-\lambda} \cdot \frac{1}{x!} I_{(0,1,\ldots)}(x) \exp(x \log \lambda)$$

Here $a(\lambda) = e^{-\lambda}$, $b(x) = \frac{1}{x!} I_{(0,1,\ldots x)}$, $c(\lambda) = \log \lambda$ and $d(x) = x$. So $p(x)$ belongs to the exponential family.

Exponential family & Sufficient Statistics.

From the definition of exponential family, from (1), we can write

$$f(\cdot, \theta) = a(\theta) \cdot b(x) \cdot \exp\left[c(\theta) \cdot d(x)\right]$$

finding the likelihood function

$$L = \prod f(\cdot, \theta) = \prod a(\theta) \cdot b(x) \exp\left[c(\theta) \cdot d(x)\right]$$

$$= a^n(\theta)\left[\prod h(x_i)\right] \exp\left[c(\theta) . \Sigma d(x_i)\right] \quad -②$$

According to the factorization theorem, T(x) is said to be sufficient estimator for the parameter θ if

20

$$L = g(T, \theta)\, h(x) . \quad -③$$

comparing ② & ③, $\Sigma d(x_i)$ is sufficient statistics for θ.

It can be shown that the sufficient statistics so obtained is minimal.

k - parameter exponential Family.

A family of densities $f(\cdot\,; \theta_1, \theta_2 \dots \theta_k)$ that can be expressed as

$$f(x; \theta_1, \theta_2 \dots \theta_k) = a(\theta_1, \theta_2 \dots \theta_k)\, b(x) \exp c_j(\theta_1, \dots \theta_k)\, d_j(\cdot)$$

for a suitable choice of functions of $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$ and is defined to belong one parameter exponential family.

Example: Consider normal distribution with μ & σ^2.

Let $\mu = \theta_1$ and $\sigma^2 = \theta_2$

$$f(x; \theta_1, \theta_2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right] . \exp\left[-\frac{1}{2\sigma^2}x^2 + \frac{\mu}{\sigma^2}x\right]$$

Here $a(\mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{1}{2}\frac{\mu^2}{\sigma^2}\right)$

$b(x) = 1$

$c_1(\mu, \sigma) = -\frac{1}{2\sigma^2}$

$c_2(\mu, \sigma) = \frac{\mu}{\sigma^2}$

$d_1(x) = x^2$

$d_2(x) = x$

∴ normal belongs to exponential family.

Example. Show that $f(x) = \frac{1}{B(\theta_1, \theta_2)} x^{\theta_1 - 1} (1-x)^{\theta_2 - 1}$, $0 < x < 1$, $\theta > 0$ belongs to exponential family