STATISTICAL INFERENCE

Inference is of two kind.
(i) Deductive → derived from known facts
(ii) Inductive → derived from some part & inferred for whole

Statistical Inference is Inductive Inference.

Parametric   Semi-parametric   Non-parametric

Here nature of pop^n is known.

THEORY OF POINT ESTIMATION

Suppose, for example, that a random variable X is known to have a normal dist^n $N(\mu, \sigma^2)$, but we donot know one of the parameter say μ.

Suppose further that a sample $(X_1, X_2, \ldots X_n)$ is taken from X.

The problem of pt. estimation is to pick a (one-dimensional) statistic $T(X_1, X_2, \ldots X_n)$ that best estimates the parameter μ.

The Statistic T is called
while numerical value of T when the realizat is $x_1, x_2, \ldots x_n$ is called

NOTE If both μ & σ^2 are unknown, we seek a joint statistic $T = (U, V)$ as an estimate of (μ, σ^2).

The Problem of Point Estimation

Let X be a r.v. defined on a probability space $(\Omega, \mathcal{S}, P)$. Suppose that the DF F of X depends on a certain number of parameters, and suppose further that the functional form of F is known except perhaps for a finite number of these parameters.

Let θ be the vector of (UNKNOWN) parameters associated with F & let θ takes values in parameter ~~space~~ set Θ

The set of all admissible values of parameters of a DF F is called the PARAMETER SPACE $\{\Theta\}$

EXAMPLE Let $X \sim B.D(n,p)$ & p be unknown, then
$\Theta = \{p : 0<p<1\}$ & $\{B(n,p) : 0<p<1\}$ is the family ~~of~~ of possible pmf's of X.

EXAMPLE : Let $X \sim N.D(\mu, \sigma^2)$. If both μ & σ^2 are unknown,
$\Theta = \{(\mu, \sigma^2) : -\infty < \mu < +\infty, \sigma^2 > 0\}$

If $\mu = \mu_0$ (say) & σ^2 is unknown, $\Theta = \{(\mu_0, \sigma^2) : \sigma^2 > 0\}$
or simply $\Theta = \{\sigma > 0\} = (0, \infty)$

DEFINITION Let $X_1, X_2, \ldots X_n$ be a sample from F_θ, where $\theta \in \Theta \subseteq \mathbb{R}$
A statistic $T(X_1, X_2, \ldots X_n)$ is said to be a (point) estimator of θ if T maps $\mathbb{R}_n$ into Θ.

Remark : Let $X_1, X_2, \ldots X_n$ be a sample from $B.D(n,p)$ where p is unknown. T can be $\overline{X}$, $\frac{1}{2}$, X_1, $\frac{X_1+X_n}{2}$,

$\therefore$ WE NEED SOME CRITERION TO CHOOSE AMONG POSSIBLE ESTIMATES

PROPERTIES OF ESTIMATES

1) UNBIASEDNESS

An estimator $T = T(x_1, x_2, \ldots x_n)$ is called an unbiased estimator of $\gamma(\theta)$ iff

$$E(T) = \gamma(\theta) \; \forall \; \theta \in \Theta,$$ where Θ is the parameter space

An estimator which is not unbiased, is called BIASED ESTIMATOR & the quantity $b(\theta) = E(T) - \gamma(\theta)$ is called BIAS of the estimator T.

The estimator T is (+)ively or (−)ively biased according as $b(\theta) > 0$ or $b(\theta) < 0$.

For a biased estimator, an estimator is called an ASYMPTOTICALLY UNBIASED of $\gamma(\theta)$ iff

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

REMARK

UNBIASNESS IS THE PROPERTY ASSOCIATED WITH FINITE SAMPLE SIZE

THEOREM If $X_1, X_2, \ldots, X_n$ are random variables with a common mean μ, then $\bar{X} = (X_1 + X_2 + \ldots + X_n)/n$ is an unbiased estimator of μ whether the X's are ind. or not, even if they have different distr's, provided only that $E(X_i) = \mu \; \forall \; i$

Proof $E(\bar{X}) = E\left[\frac{X_1 + X_2 + \ldots + X_n}{n}\right] = \frac{1}{n}\left[E(X_1) + E(X_2) + \ldots + E(X_n)\right]$

$$= \frac{1}{n}(\mu + \mu + \mu + \ldots + \mu) = \frac{n\mu}{n} = \mu$$

$\therefore \bar{X}$ is an unbiased estimator of μ. Hence Proved

THEOREM : If $X_1, X_2, \ldots X_n$ is a random sample from a distn. with finite variance $\sigma^2 > 0$ then

s^2 = Sample variance is biased

& S^2 is unbiased estimator of popn. variance σ^2.

PROOF $\sum(X_i-\mu)^2 = \sum(X_i - \bar{X} + \bar{X} - \mu)^2$

$= \sum(X_i-\bar{X})^2 + n(\bar{X}-\mu)^2 + 0 \quad (\because \sum(X_i-\bar{X})=0)$

$\therefore \quad ns^2 = \sum(X_i-\bar{X})^2 = n(\bar{X}-\mu)^2$

Taking expectation on both sides

$n E(s^2) = E\left(\sum(X_i-\mu)^2\right) - nE(\bar{X}-\mu)^2$

$\Rightarrow E(s^2) = \frac{1}{n}\sum Var(X_i) - \frac{n\, Var(\bar{X})}{n}$

$= \frac{\cancel{n}}{\cancel{n}}\sigma^2 - Var(\bar{X}) \quad \text{---} \; (*)$

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

$= \frac{1}{n^2}\sum \sigma^2 \quad (\because \text{given})$

$= \sigma^2/n$

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

$\therefore \boxed{E(s^2) \neq \sigma^2} \Rightarrow s^2$ is not unbiased estimator of σ^2

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

ie $\boxed{E(S^2) = \sigma^2} \Rightarrow S^2$ is an unbiased estimator of σ^2

Note

$\lim_{n\to\infty} E(s^2) = \sigma^2 \Rightarrow s^2$ is asymptotically unbiased estimator of σ^2

Q. Construct an example to show that an unbiased estimator may not be unique.

Ans. Let $X_1, X_2, \ldots X_n$ be a random sample from Poisson distn (α), then $\bar{X} = \frac{\Sigma X_i}{n}$ & $S^2 = \frac{1}{n-1} \Sigma(X_i - \bar{X})^2$

Since $E(X) = Var(X) = \alpha$

$\therefore$ Both $\bar{X}$ & S^2 are unbiased estimators of popn. parameter α. Thus, the unbiased estimator may not be unique.

NOTE

$$T = \frac{m\bar{x} + nS^2}{m+n}$$ is also unbiased estimator of α $\forall m, n \in \mathbb{R}$, $m \neq -n$

Q. Let $X \sim$ Bin. distn $(1, \theta)$. Find an unbiased estimator of θ^2

Ans. Let T be an unbiased estimator of θ^2

$\therefore$ $E(T) = \theta^2$, $0 \leq \theta \leq 1$

$\Rightarrow T(1) \cdot \theta + T(0) \cdot (1-\theta) = \theta^2$

($\because$ when $n=1$, Bin. distn reduces to Bernoulli distn. & $X=0$ with prob. $1-\theta$ and $X=1$, with prob θ)

Let $T(1) = a$ & $T(0) = b$

$\therefore \Rightarrow \theta^2 - (a-b)\theta - b = 0$

This quad. must hold for $\forall \theta \in [0,1]$ which is impossible

(Reason: If a convergent power series vanishes in an open interval, each of its coefficients must be zero.)

Hence

THERE IS NO UNBIASED ESTIMATOR OF θ^2.

Q. $X_1, X_2, \ldots X_n$ are random observations on a Bernoulli variate X taking the value 1 with prob. θ & value 0 with prob $(1-\theta)$.
Show that $T(T-1)/n(n-1)$ is an unbiased estimate of θ^2, where
$$T = X_1 + X_2 + \ldots + X_n$$

Soln. $T = \sum X_i \sim B.D(n, \theta)$

$$\therefore E(T) = n\theta \;,\; Var(T) = n\theta(1-\theta)$$

$$\therefore E(T^2) = Var(T) + [E(T)]^2 = n\theta(1-\theta) + n^2\theta^2$$

$$\therefore E\left[\frac{T(T-1)}{n(n-1)}\right] = \frac{1}{n(n-1)}\left[E(T^2) - E(T)\right]$$

$$= \theta^2$$

Hence Proved

Q. Let $X \sim P.D(\theta)$, Show that $(-3)^X$ is an unbiased estimator of $e^{-4\theta}$

Soln $E((-3)^X) = E(t^X) = e^{\theta(t-1)} = e^{-4\theta}$
Hence $(-3)^X$ is an unbiased estimator of $e^{-4\theta}$
However
$$T(X) = (-3)^X > 0 \quad \text{if } X \text{ is even}$$
$$< 0 \quad \text{"} \quad X \text{ is odd.}$$
but $e^{-4\theta}$ is always (+)ive

$\therefore$ An UNBIASED ESTIMATOR IS ABSURD

Q. Show that if $X \sim P.D(\theta)$ an unbiased estimator of the form aX^2+bX+C can be found for $(1+\theta)(2+\theta)$

Soln. Since $X \sim P.D(\theta) \Rightarrow E(X^{(r)}) = \theta^r$

$\therefore E(aX^2+bX+c) = aE(X^2)+bE(X)+c$
$= a\theta^2+(a+b)\theta+c.$

Setting this equal to $(1+\theta)(2+\theta)$ we get
$a=1,\ b=2 \ \& \ c=2.$

Hence X^2+2X+2 is an unbiased estimator of $(1+\theta)(2+\theta)$

Hence Proved

Q. If T is an unbiased estimator of θ. Show that $\sqrt{T}$ and T^2 are the biased estimators of $\sqrt{\theta}$ and θ^2 resp.
Show further that $2\bar{X}$ is unbiased estimate of θ for $f(x,\theta) = \frac{1}{\theta}$; $0<x<\theta$.

Soln. Given $E(T) = \theta$
We know for a non-degenerate r.v. Y
$Var(Y) \neq 0 \Rightarrow E(Y^2) \neq [E(Y)]^2$

$\therefore E(T) \neq [E(\sqrt{T})]^2 \Rightarrow \boxed{E(\sqrt{T}) \neq \sqrt{\theta}}$

$||^{ly}\ E(T^2) \neq [E(T)]^2 \Rightarrow \boxed{E(T^2) \neq \theta^2}$

Thus $\sqrt{T}$ & T^2 are biased estimators of $\sqrt{\theta}$ & θ resp.

Now if $X \sim U.D(0,1)$
$\Rightarrow E(X) = \frac{\theta}{2} \Rightarrow E(2\bar{X}) = 2E(\bar{X})$
$= 2\left[\frac{\theta}{2}\right] = \theta$

$\therefore 2\bar{X}$ is an unbiased estimator of θ for $U.D(0,\theta)$ distn.

Scanned by CamScanner

INVARIANCE PROPERTY

OF CONSISTENCY

If $T_n \xrightarrow{p} \gamma(\theta)$ and $g(\cdot)$ is a continuous f then

$$g(T_n) \xrightarrow{p} g(\gamma(\theta))$$

Proof

> Here we use a result: If $f(x)$ is cont. fn. at $x=a$ then $|f(x)-f(a)|<\epsilon$ whenever $|x-a|<\delta$ $\forall\ (\epsilon,\delta)>0$

Since $g(\cdot)$ is a cont. fn.

$\Rightarrow$ $g(T_n)$ is a cont. fn. of T_n

$$|g(T_n)-g(\gamma(\theta))|<\epsilon \text{ whenever } |T_n-\gamma(\theta)|<\delta \quad \forall\ (\epsilon,\delta)>0 \quad \text{--- (1)}$$

Since For two events A & B if $A \Rightarrow B$, then $A \subseteq B$

ie $P(A) \leq P(B)$ or $P(B) \geq P(A)$

Using above ; (1) $\Rightarrow$

$$P[\,|T_n-\gamma(\theta)|<\delta\,] \leq P[\,|g(T_n)-g(\gamma(\theta))|<\epsilon\,]$$

ie.

$$P[\,|g(T_n)-g(\gamma(\theta))|<\epsilon\,] \geq P[\,|T_n-\gamma(\theta)|<\delta\,]$$

Since $T_n \xrightarrow{p} \gamma(\theta) \Rightarrow P[\,|T_n-\gamma(\theta)|<\delta\,] \to$ as $n\to\infty$

$$\therefore \lim_{n\to\infty} P[\,|g(T_n)-g(\gamma(\theta))|<\epsilon\,] \geq 1$$

But Prob$(\cdot) \leq 1$

$$\therefore \lim_{n\to\infty} P[\,|g(T_n)-g(\gamma(\theta))|<\epsilon\,] = 1$$

ie $g(T_n) \xrightarrow{p} g(\gamma(\theta))$ Hence Proved

Remember

If $T_n \xrightarrow{p} \theta$ & $T_n' \xrightarrow{p} \theta'$

then

(i) $T_n \pm T_n' \xrightarrow{p} \theta \pm \theta'$

(ii) $T_n . T_n' \xrightarrow{p} \theta.\theta'$

(iii) $\frac{T_n}{T_n'} \xrightarrow{p} \frac{\theta}{\theta'}$ (for $\theta' > \theta$)

① Show that consistent estimator is never unique.

Sol. Let T_n be a consistent estimate of parameter θ & consider an other estimator

$$S_n = \left(\frac{n-a}{n-b}\right) T_n$$

Since $T_n \xrightarrow{p} \theta$ & $\frac{n-a}{n-b} = \frac{1-\frac{a}{n}}{1-\frac{b}{n}} \to 1$ as $n \to \infty$.

It follows that $S_n \xrightarrow{p} \theta$.

$\therefore$ S_n is also consistent estimator of same parameter θ

Hence the result

② Let $X_1, X_2, \ldots X_n$ be a random sample from a popn. whose mean is μ & variance $\sigma^2 > 0$.

Show that

$$T_n(X_1, X_2, \ldots X_n) = \frac{2}{n(n+1)} \sum_{k=1}^{n} k X_k$$ is a consistent estimator of μ

Soln $E(T_n) = \frac{2}{n(n+1)} \sum k E(X_k)$

$= \frac{2}{n(n+1)} \sum k\mu = \frac{2\mu}{n(n+1)} \left[\frac{n(n+1)}{2}\right] = \mu$

$\boxed{E(T_n) = \mu}$

& $Var(T_n) = \frac{4}{n^2(n+1)^2} \sum k^2 Var(X_k) = \frac{4\sigma^2}{n^2(n+1)^2} \sum k^2$

$= \frac{4\sigma^2}{n^2(n+1)^2} \cdot \frac{n(n+1)(2n+1)}{6} = \frac{2}{3} \sigma^2 \frac{(2n+1)}{n(n+1)} \to 0$ as $n \to \infty$

$\therefore$ By Sufficient conditions, it follows T_n is a consistent estimator of μ.

(C) Let $X_1, X_2, \ldots X_n$ be a random sample from a pop for which $(2r)^{th}$ moment about origin exist ie $\mu'_{2r} = E(X^{2r})$ exist. Show that the sample moment $m'_r = \frac{1}{n} \sum_{i=1}^{n} X_i^r$ is a consistent estimator of μ'_r.

Soln
$\therefore E(m'_r) = E\left(\frac{1}{n} \sum X_i^r\right) = \frac{1}{n} \sum E(X_i^r) = \frac{1}{n} \sum_{i=1}^{n} \mu'_r = \mu'_r \quad \forall n$

($\because$ existence of higher moments $\Rightarrow$ existence of lower moments.)

&

$Var(m'_r) = Var\left(\frac{1}{n} \sum_{i=1}^{n} X_i^r\right) = \frac{1}{n^2} \sum Var(X_i^r)$ ($\because$ Covariance terms are zero as X_i's are ind.)

$\therefore Var(m'_r) = \frac{1}{n^2} \sum \left[E(X_i^r)^2 - [E(X_i^r)]^2\right] = \frac{1}{n^2} \sum (\mu'_{2r} - (\mu'_r)^2)$

$= \mu'_{2r} - (\mu'_r)^2 \to 0$ as $n \to \infty$.

$\therefore$ Sufficient conditions are satisfied

$\Rightarrow$ m_r' is a consistent estimator of μ_r'

Hence Proved

① Let $X_1, X_2, \ldots X_n$ be a random sample of size n from a popn. with variance σ^2.

Show that
$$S^2 \xrightarrow{p} \sigma^2.$$

Soln.

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

$$= \frac{1}{n-1}\left[\sum X_i^2 + n\bar{X}^2 - 2\bar{X}\sum X_i\right]$$

$$= \frac{1}{n-1}\left[\sum X_i^2 - n\bar{X}^2\right]$$

$$= \frac{1}{n-1}\left[n m_2' - n(m_1')^2\right]$$

$$= \frac{n}{n-1}\left[m_2' - (m_1')^2\right] = \frac{n}{n-1}\sigma^2 \quad (\because \mu_2 = \sigma^2)$$

$\therefore$ $S^2 \to \sigma^2$ as $n \to \infty$.

Thus $S^2 \xrightarrow{p} \sigma^2$. Hence Proved

Efficient Estimators

Let T_1 & T_2 be 2 consistent estimators of a parameter θ.

Estimator T_1 is said to be more efficient than estimator T_2 if

$$Var(T_1) < Var(T_2) \quad \forall \; n$$

Most Efficient Estimator

A consistent estimator T of parameter θ is called Most efficient estimator if

$$Var(T) < Var(T'), \text{ where } T' \text{ is any other consistent estimator of } \theta.$$

& efficiency of T' is given by

$$E = \frac{Var(T)}{Var(T')}$$

Clearly $0 < E \leq 1$

Q. Let $X_1, X_2, \ldots, X_n$ be a random sample from normal popn. with mean μ & variance σ^2. Check the efficiency of sample mean & sample median.

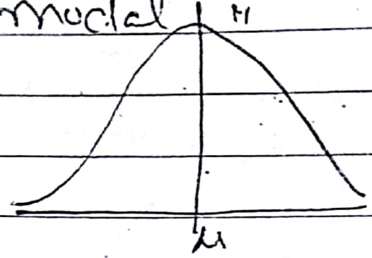
Sol. we have shown that: sample mean $(\bar{x})$ is an unbiased & consistent estimator of μ.

& From symmetry it follows that sample median (M_d) is an unbiased estimate of μ $\forall \; n$

Also for large n

$$Var(M_d) = \frac{1}{4nf_1^2}$$

where f_1 = median ordinate of the parent distn.
= modal " " " "



$$\Rightarrow f_1 = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\Big|_{x=\mu} = \frac{1}{\sigma\sqrt{2\pi}}$$

$$\therefore Var(M_d) = \frac{\pi\sigma^2}{2n}$$

$\therefore E(M_d) = \mu$ & $Var(M_d) \to 0$ as $n \to \infty$

$\therefore$ Sample mean & also sample median are consistent estimator of μ

$\therefore$ Required efficiency $(E) = \dfrac{\sigma^2/n}{\left(\frac{\sigma^2}{n}\right)\cdot\frac{\pi}{2}} = \dfrac{2}{\pi}$

$= 0.67$ Ans

($\because Var(\bar{x}) < V(M_d) \Rightarrow$ for normal distn, sample mean is more efficient estimator of μ than sample median, for large samples at least)

TRY eg. 17.7 & 17.8 p.17.8
& p.17.9 from FMS

Minimum Variance Unbiased Estimators

M.V.U.E

Let $T = T(x_1, x_2 \ldots x_n)$ based on a sample of size n

T is said to be M.V.U.E of $\gamma(\theta)$ if

(i) T is unbiased for $\gamma(\theta)$ $\forall\ \theta \in \Theta$

& (ii) $Var(T) \leq Var(T')$

where T' is any other unbiased estimator of $\gamma(\theta)$

ie T is MVUE of $\gamma(\theta)$

if

$E_\theta(T) = \gamma(\theta)\ \forall\ \theta \in \Theta$

& $Var_\theta(T) \leq Var_\theta(T')$ $\forall\ \theta \in \Theta$

where T' is any other unbiased estimator of $\gamma(\theta)$.

An M.V.U. is unique.

ie if T_1 & T_2 are M.V.U. estimators of $\gamma(\theta)$ then $T_1 = T_2$ almost surely

Soln Given $E(T_1) = E(T_2) = \gamma(\theta)\ \forall\ \theta \in \Theta$

& $Var(T_1) = Var(T_2)\ \forall\ \theta \in \Theta$

Let $T = \frac{1}{2}(T_1 + T_2)$ be a new estimator of $\gamma(\theta)$

$E(T) = \frac{1}{2}(\gamma(\theta) + \gamma(\theta)) = \gamma(\theta)$

$\therefore$ T is also unbiased estimator of $\gamma(\theta)$

$Var(T) = Var\left(\frac{1}{2}(T_1+T_2)\right)$

$= \frac{1}{4}\left[Var(T_1) + Var(T_2) + 2\,Cov(T_1, T_2)\right]$

$= \frac{1}{4}\left[Var(T_1) + Var(T_1) + 2\rho\sqrt{Var(T_1)\,Var(T_2)}\right]$

where ρ = Karl Pearson coeff of corr. bet T_1 & T_2

$\therefore Var(T) = \frac{1}{2}Var(T_1)\cdot(1+\rho)$

Since T_1 is MVU $\Rightarrow Var(T) \geq Var(T_1)$

$\Rightarrow \frac{1}{2}(1+\rho) \geq 1 \Rightarrow \boxed{\rho \geq 1}$

Since $|\rho| \leq 1 \Rightarrow \boxed{\rho = 1}$ → perfect +ve corr.

i.e. There is a linear relationship bet T_1 & T_2

let $T_1 = \alpha + \beta T_2$ ---- ⊛

$\boxed{\beta > 0}$

where α & β are constants ind. of $x_1, x_2, \ldots x_n$ but may depend on θ

ie $\alpha \equiv \alpha(\theta)$ & $\beta \equiv \beta(\theta)$

⊛ $\Rightarrow E(T_1) = \alpha + \beta(E(T_2)) \Rightarrow \theta = \alpha + \beta\theta$

$\Rightarrow \alpha = 0$ & $\beta = 1$

⊛ $\Rightarrow T_1 = T_2$

H. Proved

Scanned by CamScanner

Q. Let $X_1, X_2, \ldots X_n$ be a random sample with common mean μ & variance σ^2. Let $a_1, a_2, \ldots a_n$ be real numbers s.t $\sum_{i=1}^{n} a_i = 1$ &

let $S = \sum_{i=1}^{n} a_i X_i$

Find $E(S)$ & $\min[Var(S)]$

Also interpret the result so obtained.

Soln

$$E(S) = E\left[\sum a_i X_i\right] = \sum a_i E(X_i) = \sum a_i \mu$$
$$= \mu \sum a_i = \mu$$
$$(\because \sum a_i = 1).$$

$\therefore$ S is an unbiased estimator of μ

$$Var(S) = Var\left(\sum a_i X_i\right) = \sum a_i^2 Var(X_i) = \sigma^2 \sum a_i^2$$

$$\therefore \min Var(S) = \sigma^2 \min\left(\sum a_i^2\right)$$

let $L = \sum_{i=1}^{n} a_i^2 = \sum_{i=1}^{n-1} a_i^2 + a_n^2$

$$\therefore L = \left[a_1^2 + a_2^2 + \ldots + a_n^2\right] + \left[1 - \{a_1 + a_2 + \ldots + a_{n-1}\}\right]^2$$

$\frac{\partial L}{\partial a_1} = 0 \Rightarrow a_1 = a_n$

$\frac{\partial L}{\partial a_2} = 0 \Rightarrow a_2 = a_n$

$\frac{\partial L}{\partial a_{n-1}} = 0 \Rightarrow a_{n-1} = a_n$

$\Rightarrow$ for L to be minimum

$a_i = a_n$ $\forall$ $i = 1, 2, \ldots n-1$

Since $\sum_{i=1}^{n} a_i = 1 \Rightarrow \boxed{a_i = \frac{1}{n}}$

$$\therefore \min Var(S) = \sigma^2 \sum \frac{1}{n^2} = \frac{\sigma^2}{n}.$$

$\therefore$ OPTIMAL VALUE OF $S = \sum a_i x_i = \frac{\sum x_i}{n} = \bar{X}$

Scanned by CamScanner

SUFFICIENCY

A sufficient statistic contains all the necessary information in a sample.

defn.

Let $X_1, X_2, \ldots X_n$ be a random sample from a distn. having pdf $f(x, \theta)$, $\theta \in \Theta$.

A statistic $T = T(X_1, X_2 \ldots X_n)$ is called sufficient statistic for θ

iff

Conditional distn. of $X_1, X_2 \ldots X_n$ given $T = t$ is ind. of θ.

Mathematically

let $\underset{\sim}{X} = (X_1, X_2 \ldots X_n)$

& $\underset{\sim}{x} = (x_1, x_2 \ldots x_n)$

then T is a sufficient statistic for θ iff

$$\boxed{\text{Prob}\left[\underset{\sim}{X} = \underset{\sim}{x} \mid T(\underset{\sim}{x}) = t(\underset{\sim}{x})\right] = \underbrace{h(\underset{\sim}{x})}_{\text{ind. of } \underset{\sim}{\theta}}}$$

Q. Let $X_1, X_2 \ldots X_n$ be i.i.d Neg bin (r, p) variates (r known)

Show that $T_n = X_1 + X_2 + \ldots + X_n$ is sufficient for p.

Soln Here $f(x, p) = \binom{r+x-1}{x} p^r q^x$

$\forall X_i$ $\quad x = 0, 1, 2, \ldots$, $p + q = 1$

Consider

$$\therefore \text{Prob}\left[X_1, X_2, \dots X_n \mid T_n\right] = P\left[X_1 = x_1, X_2 = x_2 \dots X_n = x_n \mid T_n = k\right]$$

$$= \frac{\text{Prob}\left[(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) \cap (T_n = k)\right]}{P\left[T_n = k\right]}$$

$$= \frac{\text{Prob}\left[X_1 = x_1, X_2 = x_2, \dots X_{n-1} = x_{n-1}, X_n = k - (x_1 + x_2 + \dots + x_{n-1})\right]}{P\left[\sum_{i=1}^{n} X_i = k\right]}$$

$$= \frac{\prod_{i=1}^{n-1} \text{Prob}\left[X_i = x_i\right] \times \text{Prob}\left[X_n = k - x_1 - x_2 - \dots - x_{n-1}\right]}{P\left[\sum_{i=1}^{n} X_i = k\right]} \quad \text{---} \; (*)$$

Result $\quad X_i \sim \text{Neg. bin}(r, p)$

$$\sum_{i=1}^{n} X_i \sim N \cdot B(nr, p)$$

After simplification, we find
RHS of (*) is ind. of p.

$\therefore \sum X_i$ is suff. statistic of p.

Q: let $X_1, X_2 \dots X_n$ be random sample from a bernoulli distn. with parameter θ, $\theta = p$ $(0 \leq p \leq 1)$
Show that
$\sum X_i$ is suff. estimator of θ.

(take help of Illustration p. 17.15 FMS)

FACTORIZATION THEOREM
(DUE TO NEYMANN)

This theorem gives Necessary & sufficient condition for a distn. to admit sufficient statistics.

Statement: $T = t(\underset{\sim}{x})$ is sufficient estimator of θ iff.
the joint density fn. (say L) of the sample values can be expressed in the form;

$$L = g_\theta[t(\underset{\sim}{x})] * h(\underset{\sim}{x})$$

When $g_\theta(t(\underset{\sim}{x}))$ depends on θ & $\underset{\sim}{x}$ only through value of $t(x)$
& $h(\underset{\sim}{x})$ is ind. of θ

Q. Let $X_1, X_2, \ldots X_n$ be a random sample of size n from a bernoulli popn. with parameter θ.
Obtain the sufficient statistic of θ.

Soln. Joint density fn. of sample values

$= f(x_1, x_2 \ldots x_n; \theta)$

$= P_\theta(X_1 = x_1, X_2 = x_2, \ldots X_n = x_n)$

$\prod_{i=1}^{n} \text{Prob}(X_i = x_i)$

Scanned by CamScanner

$$= \prod_{i=1}^{n} \theta^{x_i} (1-\theta)^{1-x_i}$$

$$= \theta^{\sum x_i} (1-\theta)^{n-\sum x_i} \quad , \boxed{\text{let } \sum x_i = t}$$

$$= \{\theta^t (1-\theta)^{n-t}\} \cdot \{1\}$$

$$g_\theta(t(\underset{\sim}{x})) \qquad h(\underset{\sim}{x})$$

$\therefore$ By Using FACTORIZATION Th. WE CONCLUDE there exist sufficient estimator of θ.

Left part

show that $\sum_{i=1}^{n} X_i$ is suff. estimator of θ

Result

If T_n is suff. estimator for parameter θ & if $g(T_n)$ is 1-1 fn. of T_n.

then $g(T_n)$ is also suff. for θ

Note : Using above result, we conclude Sample mean is also suff. for θ

TRY: eg. 17.14, 17.15, 17.16 & 17.17 from FMS (p. 17.16 – 17.18)

Q Let $X_1, X_2, \ldots X_n$ be a r.s. from a normal popn. with mean μ & variance σ^2 find

(i) Suff. statistic for μ (σ^2 known)
(ii) " " " σ^2 (μ ")
(iii) " " " μ & σ^2 (BOTH UNKNOWN)

Soln

Joint density fn (L)

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

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2}\sum\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\left(\sum x_i^2 - 2\mu\sum x_i + n\mu^2\right)}$$

$$= \left[\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n\right]\left[e^{-\frac{1}{2\sigma^2}\left(\mu^2 - 2\mu\sum x_i\right)}\right]\left[e^{-\frac{1}{2\sigma^2}\sum x_i^2}\right] \quad -(*)$$

$$= \left[e^{-\frac{1}{2\sigma^2}\left(\mu^2 - 2\mu\sum x_i\right)}\right] \cdot \left[\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n e^{-\frac{1}{2\sigma^2}\sum x_i^2}\right]$$

$$= g\left(\sum x_i ; \mu\right) \cdot h(x_1, x_2, \ldots x_n)$$

$\therefore t = \sum_{i=1}^{n} x_i$ is suff. estimator of θ

or

$T_n = \sum x_i$ is suff estimator of θ

Also $\bar{X}$ being 1–1 f^n of T_n.

is also suff for $\underline{\theta}$.

(ii)

$$f(x_1, x_2 \ldots x_n, \sigma^2)$$

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

$$= \left[\frac{1}{\sigma^n} e^{-\frac{1}{2\sigma^2}\Sigma(x_i-\mu)^2}\right]\left(\frac{1}{\sqrt{2\pi}}\right)^n$$

$$= g\left(\Sigma(x_i-\mu)^2; \sigma^2\right) \cdot h(x_1, x_2 \ldots x_n)$$

$\therefore$ $T_n = \sum(X_i-\mu)^2$ is a suff. statics of.

$\Rightarrow$ $\underset{\wedge}{T_n} = \frac{1}{n}\Sigma(X_i-\mu)^2$ is also a " " " "

(iii) From ⊛ we have

$$f(x_1, x_2 \ldots x_n, \mu, \sigma^2)$$

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

$$= g\left(\Sigma x_i, \Sigma x_i^2, \mu, \sigma^2\right) h(x_1, x_2 \ldots x_n)$$

$\Rightarrow$ $T_n = \begin{bmatrix} \Sigma X_i \\ \Sigma X_i^2 \end{bmatrix}$ is jointly suff for $\underset{\sim}{\theta} = \begin{bmatrix} \mu \\ \sigma^2 \end{bmatrix}$

$T_n' = \begin{bmatrix} \bar{X} \\ \frac{1}{n}\Sigma(X_i-\bar{X})^2 \text{ or } \frac{\Sigma X_i^2}{n} \end{bmatrix}$ is also jointly suff for μ & $\underline{\sigma^2}$

Remark

An estimator which is unbiased & also sufficient has lower variance than that which is only unbiased.

RAO - BLACKWELL THEOREM

Let $X_1, X_2, \ldots X_n$ be a random sample from a popn. with density $f(x; \theta)$ & let S_1 be sufficient estimator for θ. Let the statistic T_n be unbiased for $\gamma(\theta)$, $\theta \in \Theta$

Define $T_n' = E(T_n | S_1)$

then

(i) T_n' is a statistic

(ii) T_n' " " fn. of sufficient statistic S_1.

(iii) $Var(T_n') \leq Var(T_n)$; $\theta \in \Theta$ with strict inequality sign holding for atleast one θ.

Remark If the estimator T_n is not unbiased estimator of θ then we judge its merits & make efficiency comparisons on the basis of Mean Square Error (MSE) defined as

$$MSE(T_n) = E(T_n - \theta)^2$$

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

$$MSE(T_n) = E\left(T_n - E(T_n)\right)^2 + [E(T_n) - \theta]^2$$
$$+ 2E\left[\left(T_n - E(T_n)\right)\underbrace{\left(E(T_n) - \theta\right)}_{\text{constant}}\right]$$
$$= Var(T_n) + [E(T_n) - \theta]^2 + 0$$
$$\therefore MSE(T_n) = Var(T_n) + [BIAS]^2.$$

Q Let $X_1, X_2, \ldots X_n$ be a random sample from Uniform popn. on $(0, \theta)$. Find sufficient estimator for θ.

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

Let

$t = \max(x_1, x_2, \ldots x_n) = x_{(n)}$.

then pdf of t ie $x_{(n)}$

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

Here $F(x) = P[X \leq x] = \int_0^x f(x, \theta)\, dx = \frac{x}{\theta}$

$\therefore g(t, \theta) = \frac{n}{\theta^n}[x_{(n)}]^{n-1}$.

$\Rightarrow \frac{1}{\theta^n} = \frac{g(t, \theta)}{n[x_{(n)}]^{n-1}}$

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

$= g(t, \theta) \cdot h(x_1, x_2, \ldots x_n)$

$\therefore$ By FISHER-NEYMANN CRITERION

- - ---- Statistic $t = x_{(n)}$ is sufficient estimator for θ

i.e $T = X_{(n)}$ is sufficient

Complete family of DISTRIBUTIONS

A statistic $T = t(x)$ is said to be complete for $\theta \in \Theta$ if

$E_\theta[h(T)] = 0 \; \forall \theta$

$$\Rightarrow P_\theta[h(T) = 0] = 1$$

i.e $\int h(t) g(t,\theta)\, dt = 0 \quad \forall \theta \in \Theta$

or $\sum h(t) g(t,\theta) = 0 \quad \forall \theta \in \Theta$

$$\Rightarrow h(T) = 0 \; \forall \theta \in \Theta$$

almost surely (a.s.)

Question cont.

$T = X_{(n)}$ has pdf

$$g(t,\theta) = \begin{cases} \dfrac{n t^{n-1}}{\theta^n} & 0 \le t \le \theta \\ 0 & \text{otherwise} \end{cases}$$

$\therefore E_\theta[h(T)] = 0 \Rightarrow \dfrac{n}{\theta^n} \int_0^\theta h(u) \cdot u^{n-1}\, du = 0$
$\forall \theta \in \Theta$ $\qquad \forall \theta \in \Theta$

Diff w.r.t θ, we get

$h(\theta) \cdot \theta^{n-1} = 0 \quad \forall \ \theta \in \Theta$

$\Rightarrow \boxed{h(T) = 0 \ , a.s.}$

$\therefore$ $T = X_{(n)}$ is also complete for θ.

ie $T = X_{(n)}$ is complete suff. statistic for θ

Now

$$E(T) = E(X_{(n)}) = \left(\frac{n}{n+1}\right)\theta.$$

$$\Rightarrow E\left[\frac{(n+1)T}{n}\right] = \theta$$

ie $T_n = \frac{(n+1)X_{(n)}}{n}$ is unbiased for θ

THEOREM: Let T be a complete suff. statistic for $\gamma(\theta)$, $\theta \in \Theta$. Then $\phi(T) = E(U \mid T=t)$ where U is an unbiased estimator of $\gamma(\theta)$ is the unique unbiased estimator for $\gamma(\theta)$

Using above result we conclude

$T_n = \frac{n+1}{n} X_{(n)}$ is MVU estimator of θ