LECTURE 10

SOME PROPERTIES OF ESTIMATORS

ESTIMATOR

Assume that we have a sample ($x_1, x_2, \ldots, x_n$) from a given population. All parameters of the population are known except some parameter θ. We want to determine from the given observations unknown parameter - θ. In other words we want to determine a number or range of numbers from the observations that can be taken as a value of θ.

Estimator – is a method of estimation.

Estimate – is a result of an estimator

Point estimation – as the name suggests is the estimation of the population parameter with one number.

PROPERTIES OF ESTIMATORS

Since estimator gives rise an estimate that depends on sample points $(x_1,x_2,\ldots,x_n)$ estimate is a function of sample points. Sample points are random variable therefore estimate is random variable and has probability distribution. We want that estimator to have several desirable properties.

Goal:

– Check how *good* are these estimator(s). Or are they *good* at all?

– If more than one *good* estimator is available, which one is *better*?

1. UNBIASEDNESS

If an estimator $\hat{\theta}$ estimates θ then difference between them ($\hat{\theta}$ - θ) is called the estimation error. Bias of the estimator is defined as the expectation value of this difference

$$Bias_{\theta} = E(\hat{\theta} - \theta) = E(\hat{\theta}) - \theta$$

If the bias is equal to zero then the estimation is called unbiased.

Therefore, $E\left[\hat{\theta}\right] = \theta \text{ for all } \theta \in \Omega$

Example

The sample mean is an unbiased estimator:

$$E(\bar{x}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x) = \frac{1}{n} n\mu = \mu$$

$$\text{Bias}_{\mu} = \text{E}(\bar{x}\text{-}\mu) = 0$$

Here we used the fact that expectation and summation can change order and the expectation of each sample point is equal to the population mean.

Example

Given sample of size n from the population with unknown mean (θ) and variance (σ^2) we estimate mean as we already know and variance (intuitively) as:

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

What is the bias of this estimator?

$$E(\hat{\sigma}^2) = \frac{1}{n}\sum_{i=1}^{n}E(x_i^2) - E((\frac{1}{n}\sum_{i=1}^{n}x_i))^2 = E(x^2) - \frac{1}{n^2}E(\sum_{i=1,\, j=1}^{n}x_i x_j)$$

$$= E(x^2) - \frac{1}{n^2}(E(\sum_{i=1}^{n}x_i^2) - E(\sum_{i\neq j}^{n}x_i x_j)) = E(x^2) - \frac{1}{n^2}(nE(x^2) - n(n-1)E(x)^2)$$

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

- Sample variance is not an unbiased estimator for the population variance. That is why when mean and variance are unknown the following equation is used for sample variance:

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

1.2 ASYMPTOTICALLY UNBIASEDNESS

- An estimator $\hat{\theta}$ is an Asymptotically unbiased of the unknown parameter θ, if

$$\underset{\theta}{Bias}\left(\hat{\theta}\right) \neq 0 \; but \; \lim_{n\to\infty} \underset{\theta}{Bias}\left(\hat{\theta}\right) = 0$$

Example

Given sample of size n from the population remember the biased estimator of sample variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \qquad E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

As $n \to \infty$ $\frac{n-1}{n} \to 1$ then, $E(\hat{\sigma}^2) \to \sigma^2$

$\hat{\sigma}^2$ is a asymptotically unbiased estimator of population variance.

2. CONSISTENCY

We would like that estimator stays as close as possible to the parameter it estimates as sample size increases.

An estimator $\hat{\theta}$ which converges in probability to an unknown parameter θ for all $\theta \in \Omega$ is called a consistent estimator of θ.

$$\hat{\theta} \xrightarrow{p} \theta.$$

The property of consistency is a limiting property. It does not require any behaviour of the estimator for a finite sample size.

If there is one consistent estimator then you can construct infinitely many others. For example if $\hat{\theta}$ is consistent then $\hat{\theta}\, n/(n-1)$ is also consistent.

Example

For a r.s. of size *n* then, we know that sample mean is an unbiased estimator (UE) of population mean,

$$E(\bar{X}) = \mu \Rightarrow \bar{X} \text{ is an UE of } \mu.$$

In addition from WLLN,

$$\bar{X} \xrightarrow{p} \mu$$

As a result,

$$\Rightarrow \bar{X} \text{ is a consistent estimator (CE) of } \mu.$$

3. MEAN SQUARE ERROR

Expectation value of the square of the differences between estimator and the expectation of the estimator is called its variance:

$$Var_{\theta} = E(\hat{\theta} - E(\hat{\theta}))^2$$

The difference between estimator and the parameter is error of the estimation. Expectation value of this error is bias. Expectation value of square of this error is called mean square error (mse).

$$MSE_{\theta} = E(\hat{\theta} - \theta)^2$$

It can be expressed by the bias and the variance of the estimator:

$$MSE_\theta(\hat{\theta}) = E(\hat{\theta}-\theta)^2 = E(\hat{\theta}-E(\hat{\theta})+E(\hat{\theta})-\theta)^2 =$$

$$E(\hat{\theta}-E(\hat{\theta}))^2 + (E(\hat{\theta})-\theta)^2 = Var_\theta(\hat{\theta}) + Bias_\theta^2(\hat{\theta})$$

MSE is equal to square of the estimator's bias plus variance of the estimator. If the bias is 0 then MSE is equal to the variance. In ideal world we would like to have minimum variance unbiased estimator.

If $\underset{\theta}{MSE}\left(\hat{\theta}\right)$ is smaller, $\hat{\theta}$ is a better estimator of θ.

For two estimators, $\hat{\theta}_1$ *and* $\hat{\theta}_2$ *of* θ*, if*

$\underset{\theta}{MSE}\left(\hat{\theta}_1\right) < \underset{\theta}{MSE}\left(\hat{\theta}_2\right), \theta \in \Omega \quad \hat{\theta}_1$ *is better estimator of* θ.

Example

- If X1, . . . , Xn ~ Uni(0, θ), To find the MSE of an estimator of population mean, we need the mean and variance of sample mean. We know that

$E[X_i] = \theta/2$ and $Var[X_i] = \theta^2/12$.

- Then,

$$E(\overline{x}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\frac{n\theta}{2} = \frac{\theta}{2}$$

$$Var(\overline{x}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\frac{n\theta^2}{12} = \frac{\theta^2}{12n}$$

$$MSE(\overline{x}) = \frac{\theta^2}{12n} + \left(\frac{\theta}{2} - \theta\right)^2 = \frac{(3n+1)\theta^2}{12n}$$

- Suppose we have one more estimator $\phi(\mathrm{x})=2\overline{x}$
- Then,

$$E(\phi(\mathrm{x})) = 2E(\overline{x}) = \theta$$

$$Var(\phi(\mathrm{x})) = 4Var(\overline{x}) = \frac{\theta^2}{3n}$$

$$MSE(\phi(\mathrm{x})) = \frac{\theta^2}{3n} + \left(\theta - \theta\right)^2 = \frac{\theta^2}{3n}$$

- As a result $\phi(\mathrm{x})=2\overline{x}$ has a smaller MSE than $\overline{x}$

4. MEAN SQUARED ERROR CONSISTENCY

If we collect a large number of observations, we hope we have a lot of information about any unknown parameter θ, and thus we hope we can construct an estimator with a very small MSE. $\hat{\theta}$ is called mean squared error consistent if

$$MSE_{\theta} = E(\hat{\theta} - \theta)^2 \to 0 \quad as \quad n \to \infty$$

Theorem: $\hat{\theta}$ is consistent in MSE iff

$i)$ $Var(\hat{\theta}) \to 0$ as $n \to \infty$.

$ii)$ $\lim_{n\to\infty} E\left[\hat{\theta}\right] = \theta.$

Example

- If X1, . . . , Xn ∼ Uni(0, θ), We know that

$$E(\overline{x}) = E(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = \frac{1}{n}\frac{n\theta}{2} = \frac{\theta}{2}$$

$$Var(\overline{x}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\frac{n\theta^2}{12} = \frac{\theta^2}{12n}$$

$$MSE(\overline{x}) = \frac{\theta^2}{12n} + \left(\frac{\theta}{2} - \theta\right)^2 = \frac{(3n+1)\theta^2}{12n}$$

- Then,

$$\lim_{n\to\infty} MSE(\overline{x}) = \lim_{n\to\infty}\frac{(3n+1)\theta^2}{12n} = \frac{\theta^2}{4} \neq 0$$

- Suppose we have one more estimator $\phi(\mathrm{x})=2\overline{x}$ we know that

$$E(\phi(\mathrm{x})) = 2E(\overline{x}) = \theta$$

$$Var(\phi(\mathrm{x})) = 4Var(\overline{x}) = \frac{\theta^2}{3n}$$

$$MSE(\phi(\mathrm{x})) = \frac{\theta^2}{3n} + \left(\theta - \theta\right)^2 = \frac{\theta^2}{3n}$$

$$\lim_{n\to\infty} MSE(\phi(\mathrm{x})) = \lim_{n\to\infty} \frac{\theta^2}{3n} = 0$$

- As a result $\phi(\mathrm{x})=2\overline{x}$ is consistent in MSE but $\overline{x}$ is not.

Example

$X \sim Exp(\theta)$, $\theta > 0$. For a r.s of size n, consider the following estimators of θ, and discuss their bias and consistency.

$$T_1 = \frac{\sum_{i=1}^{n} X_i}{n}, \quad T_2 = \frac{\sum_{i=1}^{n} X_i}{n+1}$$

Which estimator is better?

5. EFFICIENCY

- Assume we have two *unbiased* estimators of θ, i.e.

$$\hat{\theta}^{(1)}, \hat{\theta}^{(2)} : E\left(\hat{\theta}^{(1)}\right) = E\left(\hat{\theta}^{(2)}\right) = \theta$$

- If $Var\left(\hat{\theta}^{(1)}\right) \leq Var\left(\hat{\theta}^{(2)}\right)$ with strict inequality than $\hat{\theta}^{(1)}$ is more efficient than $\hat{\theta}^{(2)}$
- The *efficiency* of an unbiased estimator is defined as:

$$eff\left(\hat{\theta}^{(j)}\right) = \frac{\min_i \left\{Var\left(\hat{\theta}^{(i)}\right)\right\}}{Var\left(\hat{\theta}^{(j)}\right)} \leq 1$$

- An estimator is said to be efficient if in the class of unbiased estimators it has minimum variance.

Example

- Let $\hat{\mu}^{(1)} = \bar{X} = \frac{1}{n}\sum_{1}^{n} X_i \; ; n > 2 \quad \text{and} \quad \hat{\mu}^{(2)} = \frac{X_1 + X_n}{2}$
- Then $E\left(\hat{\mu}^{(1)}\right) = \mu \; ; E\left(\hat{\mu}^{(2)}\right) = \frac{E\left(X_1\right) + E\left(X_2\right)}{2} = \frac{\mu + \mu}{2} = \mu$
- Both estimators are unbiased.

$$Var\left(\hat{\mu}^{(1)}\right) = \left(\frac{1}{n}\right)^2 \sum_{1}^{n} Var\left(X_i\right) = \frac{1}{n^2} \cdot n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

$$Var\left(\hat{\mu}^{(2)}\right) = \frac{1}{4} \cdot \left(Var\left(X_1\right) + Var\left(X_2\right)\right) = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2} > \frac{\sigma^2}{n}$$

$\hat{\mu}^{(1)}$ is more efficient than $\hat{\mu}^{(2)}$

6. SUFFICIENCY

- An estimator is sufficient if it uses all the sample information.
- A function *U* of the sample values of a sample *x*, i.e.

  $Y=U(X_1, X_2,\ldots,X_n)$ is a *statistic* that is *sufficient* for the parameter θ if the conditional distribution of the sample random variables $h(x_1,x_2,\ldots,x_n|y)$ does not depend on θ, i.e.

- What does it mean in practice?
  - If *Y* is sufficient for θ then no more information about θ than what is contained in *Y* can be obtained from the sample.
  - It is enough to work with *Y* when deriving point estimates of θ

- The conditional distribution of sample rvs given the value of *y* of *Y*, is defined as

$$h\left(x_1, x_2, \cdots, x_n \middle| y\right) = \frac{f\left(x_1, x_2, \cdots, x_n, y; \theta\right)}{g\left(y; \theta\right)}$$

$$h\left(x_1, x_2, \cdots, x_n \middle| y\right) = \frac{L\left(\theta; x_1, x_2, \cdots, x_n\right)}{g\left(y; \theta\right)}$$

- If *Y* is a ss for θ, then

Not depend on θ for every given *y*.

$$\underbrace{h\left(x_1, x_2, \cdots, x_n \middle| y\right)}_{\text{ss for } \theta} = \frac{L\left(\theta; x_1, x_2, \cdots, x_n\right)}{g\left(y; \theta\right)} = H\left(x_1, x_2, \cdots, x_n\right)$$

may include *y* or constant.

Example

Assume $\boldsymbol{x} = (x_1, x_2)$ is a sample from $Exp(\mu)$

Let $\mathrm{U}(x_1, x_2) = x_1 + x_2$ and assume $\mathrm{U} = u$ is observed

$\Rightarrow x_2 = u - x_1.$

$$f_{X_1, X_2, U}(y_1, y_2, u) = f_{X_1, X_2}(y_1, u - y_1) =$$

$$\mu^{-1} e^{-\theta^{-1} y_1} \cdot \mu^{-1} e^{-\mu^{-1}(u - y_1)} = \mu^{-2} e^{-\mu^{-1} u}$$

$f_U(u)$?

Derive by differentiating $F_U(u) = \Pr(U \leq u)$

$$\Pr\left(U \le u\right) = \Pr\left(X_1 + X_2 \le u\right) = \Pr\left(X_2 \le u - X_1\right) =$$

$$= \iint\limits_{y_2 \le u - y_1} \mu^{-1} e^{-\theta^{-1} y_1} \cdot \mu^{-1} e^{-\mu^{-1} y_2} dy_1 dy_2 = \int_{y_1=0}^{u} \int_{y_2=0}^{u-y_1} \mu^{-1} e^{-\mu^{-1} y_1} \cdot \mu^{-1} e^{-\mu^{-1} y_2} dy_2 dy_1$$

$$= \int_{y_1=0}^{u} \mu^{-1} e^{-\theta^{-1} y_1} \int_{y_2=0}^{u-y_1} \mu^{-1} e^{-\mu^{-1} y_2} dy_2 dy_1 = \int_{y_1=0}^{u} \mu^{-1} e^{-\mu^{-1} y_1} \cdot \left[-e^{-\mu^{-1} y_2}\right]_{y_2=0}^{u-y_1} dy_1$$

$$= \int_{y_1=0}^{u} \mu^{-1} e^{-\mu^{-1} y_1} \cdot \left(1 - e^{-\mu^{-1}(u-y_1)}\right) dy_1 = \int_{y_1=0}^{u} \left(\mu^{-1} e^{-\mu^{-1} y_1} - \mu^{-1} e^{-\mu^{-1} u}\right) dy_1$$

$$= \left[-e^{-\mu^{-1} y_1} - y_1 \mu^{-1} e^{-\mu^{-1} u}\right]_{y_1=0}^{u} = -e^{-\mu^{-1} u} - t \mu^{-1} e^{-\mu^{-1} u} + 1 + 0$$

$$= 1 - e^{-\mu^{-1} u} - u \mu^{-1} e^{-\mu^{-1} u}$$

$$f_U\left(u\right) = \mu^{-1} e^{\mu^{-1} u} - \left(\mu^{-1} e^{-\mu^{-1} u} - u \mu^{-2} e^{\mu^{-1} u}\right) = u \mu^{-2} e^{\mu^{-1} u}$$

$$f_{X_1, X_2 U}\left(y_1, y_2 \middle| U = u\right) = \frac{\mu^{-2} e^{\mu^{-1} u}}{u \mu^{-2} e^{\mu^{-1} u}} = \frac{1}{u} \ \text{ not depending on } \ \mu$$

NEYMAN'S FACTORIZATION THEOREM

- *Y* is a sufficient statistic for θ iff the likelihood function can be written

$$L(\theta) = k_1(y;\theta)\,k_2(x_1, x_2, \cdots, x_n)$$

The likelihood function

Does not depend on x_i except through y

Not depend on θ (also in the range of x_i.)

where k_1 and k_2 are non-negative functions.

ANCILLARY STATISTIC

- A statistic $S(X)$ whose distribution does not depend on the parameter θ is called an **ancillary statistic**.

- Unlike a ss, an ancillary statistic contains no information about θ.

Example

Let Xi~Unif(θ,θ+1) for i=1,2,…,n

The joint pdf and cdf

$$f(x|\theta)=\begin{cases}1 & \theta<x_i<\theta+1\\ 0 & otherwise\end{cases}$$

$$F(x|\theta)=\begin{cases}0 & x_i\le\theta\\ x-\theta & \theta<x_i<\theta+1\\ 1 & x_i\ge\theta+1\end{cases}$$

The joint pdf of $X_{(1)}$ and $X_{(n)}$

$$g(x_{(1)},x_{(n)}|\theta)=\begin{cases}n(n-1)(x_{(n)}-x_{(1)})^{n-2} & \theta<x_{(1)}<x_{(2)}<\theta+1\\ 0 & otherwise\end{cases}$$

Then, the pdf for range R=X(n)-X(1) is

$$h((x_{(1)}-x_{(n)})|\theta)=\begin{cases}n(n-1)(x_{(n)}-x_{(1)})^{n-2}(1-(x_{(n)}-x_{(1)})) & 0<(x_{(n)}-x_{(1)})<1\\ 0 & otherwise\end{cases}$$

As a result R=X(n)-X(1) is an ancillary statistic because its pdf does not depend on θ.

Example

- Suppose $X \in \mathrm{Exp}(\mu)$

$$\text{Let } \mathrm{U}(\boldsymbol{x}) = \sum_{1}^{n} x_i$$

$$L(\mu; \boldsymbol{x}) = \prod_{1}^{n} f(x_i; \theta) = \prod_{1}^{n} (1/\mu) e^{-x_i/\mu} = \frac{1}{\mu^n} e^{-\frac{1}{\mu}\sum_{1}^{n} x_i} =$$

$$= \underbrace{\frac{1}{\mu^n} e^{-\frac{1}{\mu}\sum_{1}^{n} x_i}}_{K_1\left(\sum_{1}^{n} x_i; \mu\right)} \cdot \underbrace{1}_{K_2(\boldsymbol{x})} \Rightarrow \sum_{1}^{n} x_i \text{ is sufficient for } \mu$$

7. MINIMAL SUFFICIENT STATISTICS

- A ss $U(X)$ is called minimal ss if, for any other ss $U'(X)$, $U(x)$ is a function of $U'(x)$.

- **Theorem:** Let $f(x;\theta)$ be the pmf or pdf of a sample $X_1, X_2,\ldots,X_n$. Suppose there exist a function $U(x)$ such that, for two sample points $x_1,x_2,\ldots,x_n$ and $y_1,y_2,\ldots,y_n$, the ratio

$$\frac{f\left(x_1, x_2, \cdots, x_n; \theta\right)}{f\left(y_1, y_2, \cdots, y_n; \theta\right)}$$

is constant with respect to θ iff $U(x)=U(y)$. Then, $U(X)$ is a minimal sufficient statistic for θ.

RAO-BLACKWELL THEOREM

- Let $\hat{\theta}$ be an estimator of θ with $\mathrm{E}\,(\hat{\theta}^2) < \infty$ for all θ. Suppose that T is sufficient for θ, and let

$$\theta^* = \mathrm{E}\,(\hat{\theta} \mid \mathrm{T})$$

- Then for all θ,

$$\mathrm{E}\,(\theta^* - \theta)^2 \leq \mathrm{E}\,(\hat{\theta} - \theta)^2$$

- The inequality is strict unless $\hat{\theta}$ is a function of T.
- The following theorem says that if we want an estimator with small MSE we can confine our search to estimators which are functions of the sufficient statistic

Example

- Suppose $X_1, X_2, \ldots, X_n \sim Poisson(\lambda)$

$$L(x;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i\,!} = \frac{\left(\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}\right)}{\prod_{i=1}^{n} x_i\,!}$$

- Then $t = \sum_{i=1}^{n} x_i$ is an sufficient statistic.
- Suppose we have an unbiased estimator $\hat{\lambda} = X_1$ from Rao-Blackwell Theorem

$$\lambda^* = \mathrm{E}\,(\hat{\lambda} = X_1 \mid t = \sum_{i=1}^{n} x_i)$$

$$\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{E}\,(\mathrm{x}_i \mid t = \sum_{i=1}^{n} x_i) = \mathrm{E}\left(\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x}_i \mid t = \sum_{i=1}^{n} x_i\right) = t$$

- By the fact thatX1,... ,Xn are IID, every term within the sum on the l.h.s. must be the same, and hence equal to t/n. $\overline{x} = \lambda^* = \hat{\lambda}$

8. THE MINIMUM VARIANCE UNBIASED ESTIMATOR

- **Rao-Blackwell Theorem:** If $\hat{\theta}$ is an unbiased estimator of θ, and T is a ss for θ, then $\theta^* = \mathrm{E}(\hat{\theta} \mid \mathrm{T})$ is
  - an UE of θ, i.e., $E(\theta^*) = \mathrm{E}(\mathrm{E}(\hat{\theta} \mid \mathrm{T})) = \mathrm{E}(\hat{\theta}) = \theta$ and
  - the MVUE of θ.

9. COMPLETENESS

- Let {f(x; θ), θ∈Ω} be a family of pdfs (or pmfs) and U(x) (sufficient) be an arbitrary function of *x* not depending on θ. If
$$E_\theta\left[U(X)\right]=0 \text{ for all } \theta\in\Omega$$
requires that the function itself equal to 0 for all possible values of x; then we say that this family is a complete family of pdfs (or pmfs).
$$E_\theta\left[U(X)\right]=0 \text{ for all } \theta\in\Omega \Rightarrow U(x)=0 \text{ for all } x.$$

i.e., the only unbiased estimator of 0 is 0 itself.

Example

- Suppose $X_1, X_2, \ldots, X_n \sim Poisson(\lambda)$
- The statistic $y = \sum_{i=1}^{n} x_i$ has a pdf $f(y;\lambda) = \dfrac{(n\lambda)^y e^{-n\lambda}}{y!}$
- Suppose there is u(y) any fnc of the sufficient statistic y.

$$E(u(y)) = \sum_{y=0}^{\infty} u(y) \frac{(n\lambda)^y e^{-n\lambda}}{y!} = e^{-n\lambda}\left(u(0) + u(1)\frac{(n\lambda)}{1!} + \ldots \right)$$

- Since $e^{-n\lambda}$ can not be 0 E(u(y))=0 if

$$\left(u(0) + u(1)\frac{(n\lambda)}{1!} + \ldots \right) = 0$$

- However, if such an infinite series converges to zero for all λ> 0, then each of the coefficients must equal zero. Then, u(0)=u(1)=u(2)=…=0. Then the family $f(y;\lambda)$ is complete.

10. COMPLETE AND SUFFICIENT STATISTICS

- *Y* is a complete and sufficient statistic (css) for θ if *Y* is a ss for θ and the family

$$f(y;\lambda) \quad \boxed{\text{The pdf of } Y.}$$

is complete.

1) *Y* is a ss for θ.

2) u(Y) is an arbitrary function of Y.
E(u(Y))=0 for all $\theta \in \Omega$ implies that u(y)=0 for all possible Y=y.

BASU THEOREM

- If *T(X)* is a complete and minimal sufficient statistic, then *T(X)* is independent of every ancillary statistic.
- In other words in a complete family, every ancillary statistic is independent of the minimal sufficient statistic.
- Basu's Theorem is useful for deducing independence of two statistics:
- No need to determine their joint distribution
- Needs showing completeness

Example

- Let T=X1+ X2 and U=X1 - X2
- X1, X2~N(μ,σ^2), independent, σ^2 known.
- We know that T is a complete minimal ss.
- U~N(0, σ^2) $\rightarrow$ distribution free of μ

$\rightarrow$ T and U are independent by Basu's Theorem

LEHMANN-SCHEFFE THEOREM

- Let Y be a css for θ. If there is a function Y which is an UE of θ, then the function is the Uniform Minimum Variance Unbiased Estimator (UMVUE) of θ.
- Y css for θ.
- $T(y)=fn(y)$ and $E[T(Y)]=\theta$.
- $T(Y)$ is the UMVUE of θ.
- So, it is the best estimator of θ.

Example

- Suppose $X_1, X_2, \ldots, X_n \sim Poisson(\lambda)$
- The statistic $y = \sum_{i=1}^{n} x_i$ has a pdf $f(y;\lambda) = \dfrac{(n\lambda)^y e^{-n\lambda}}{y!}$

$$L(x;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\left(\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}\right)}{\prod_{i=1}^{n} x_i!}$$

- Suppose there is u(y) any fnc of the sufficient statistic y.

$$E(u(y)) = \sum_{y=0}^{\infty} u(y) \frac{(n\lambda)^y e^{-n\lambda}}{y!} = e^{-n\lambda}\left(u(0) + u(1)\frac{(n\lambda)}{1!} + \ldots\right) = 0 \quad if \quad \mathrm{u(0)=u(1)=u(2)=\ldots=0.}$$

- Then the family $f(y;\lambda)$ is complete. As a result y is css. Suppose $\bar{x} = \hat{\lambda}$ $E(\bar{x} = \hat{\lambda}) = E(\frac{1}{n}\sum_{i=1}^{n} x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n} E(x) = \frac{1}{n} n\lambda = \lambda$
- Then since $\hat{\lambda}$ is unbiased estimator of λ and a function of css y then it is UMVUE (the best estimator).

Note

- The estimator found by Rao-Blackwell Thm may not be unique. But, the estimator found by Lehmann-Scheffe Thm is unique.