Basic Statistics

Dr. Sudipta Das

Department of Computer Science,
Ramakrishna Mission Vivekananda Educational & Research Institute

Outline I

1. Statistical Inference
   - Introduction
   - Point Estimation

Chapter 8: Statistical Inference

Statistical Inference I

- The main objective in any statistical enquiry is the properties of one or more population.
  - However, the population(s) is (are) usually unknown to us, and we simply have a sample from the population (or, a sample from each of the given populations)

Statistical Inference II

- Statistical Inference:-
  Given the properties of the sample (or, of the samples), to infer about those of the population(s) is the problem of statistical inference
  - It is analogous to the inductive logic, the only difference being that the induction is achieved under probabilistic framework
    - Probability comes due to random sampling
  - It is a process of going over from the known sample to unknown population.

Statistical set-up of the problem of inference

- Let $(X_1, X_2, \ldots, X_n)$ be a random sample of size n drawn from a population (discrete/continuous) with p.m.f/p.d.f $f(\underline{x}; \underline{\theta}) = f_{\underline{\theta}}(\underline{x})$, where $\underline{\theta}$ is the unknown parameter(s) of interest.
    - Our problem is to infer about $\underline{\theta}$
- Let Θ be the set of all possible values of θ
    - Θ is called the parameter space
- Note:
    - In the problem of statistical inference, Θ is known, although θ is unknown.
    - Example 1: $X_1, X_2, \ldots, X_n \sim$ Bernoulli(p)
        - $\theta = p$ is unknown, $\Theta = [0, 1]$ is known
    - Example 2: $X_1, X_2, \ldots, X_n \sim$ Normal(μ, σ)
        - $\underline{\theta} = [\mu, \sigma]'$ is unknown, $\Theta = (-\infty, \infty) \times (0, \infty)$ is known
        - $\theta = \mu$ is unknown, $\Theta = (-\infty, \infty)$ is known
        - $\theta = \sigma$ is unknown, $\Theta = (0, \infty)$ is known

- Statistical Inference
  1. Estimation
     i. Point Estimation
     ii. Interval Estimation
  2. Hypothesis-testing

1. Estimation:-

    Here, we have **no idea** about the true value of θ and the problem is **to estimate** the likely value of θ on the basis of the random sample $(X_1, X_2, \ldots, X_n)$ drawn from the population i Point Estimation:-

Here, we estimate θ by a **single** value (i.e., by a point)

- Let $T = T(X_1, X_2, \ldots, X_n)$ be a statistic which is used to estimate the parameter θ, is called an **estimator** of θ
- For the observed sample $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, the observed value of the estimator, namely,

$$t = T(x_1, x_2, \ldots, x_n)$$

is called an estimate of θ

ii Interval Estimation:-

Here, we estimate θ by an **interval** of values

- Let $T_1 = T_1(X_1, X_2, \ldots, X_n)$ and $T_2 = T_2(X_1, X_2, \ldots, X_n)$ be two statistics such that

$$P[T_1 \leq \theta \leq T_2] = 1 - \alpha,$$

where α is a pre-assigned small quantity. Usually, we take $\alpha = 0.05$ or 0.01 etc.

- If $\alpha = 0.05$, then $P[T_1 \leq \theta \leq T_2] = 0.95$. Hence, the observed values of $[T_1, T_2] = [t_1, t_2]$, say, is called a 95% confidence interval of θ

2 Hypothesis-testing:-
Here we have some idea about the true value of θ, in the form of a hypothesis, say, $\theta = \theta_0$,

- Our problem is **to judge or test** the validity/ feasibility/ tenability of the given hypothesis $\theta = \theta_0$ on the basis of random sample of the population

Chapter 8a: Point Estimation

Point Estimation I

- In this case, we estimate θ by a **single** value (i.e., by a point)
  - Let $T = T(X_1, X_2, \ldots, X_n)$ be a statistic which is used to estimate the parameter θ, is called an **estimator** of θ
  - For the observed sample $(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n)$, the observed value of the estimator, namely,

$$t = T(x_1, x_2, \ldots, x_n)$$

is called an **estimate** of θ.

Point Estimation II

- Methods of calculating point estimates
  - Method of moments
  - Method of maximum likelihood

- Method of moments:
  It consists in equating the first few moments of the population $(\mu'_k = E[X^k])$ with the corresponding moments of the sample $\left(m'_k = \frac{1}{n}\sum_i^n x_i\right)$, i.e.,

$$\mu'_k = m'_k.$$

Point Estimation IV

- The method of moments procedure:
  Suppose there are l parameters to be estimated, say $\theta = (\theta_1, \ldots, \theta_l)'$.
  - Find l population moments, μ'_k, for $k = 1, 2, \ldots, l$.
    - μ'_k will contain one or more parameters $\theta_1, \ldots, \theta_l$.
  - Find the corresponding l sample moments, m'_k, for $k = 1, 2, \ldots, l$.
    - The number of sample moments should equal the number of parameters to be estimated.
  - From the system of equations, $\mu'_k = m'_k$, for $k = 1, 2, \ldots, l$, solve for the parameter $\theta = (\theta_1, \ldots, \theta_l)'$;
    - This will be a moment estimator of $\hat{\theta}$

Point Estimation (method of moment)

To Estimate	Notation	Point
Mean	$\bar{X}$	$\frac{1}{n}\sum_{i=1}^{n} X_i$
Proportion	$\hat{p}$	$\frac{1}{n}\sum_{i=1}^{n} I_{(X_i=1)}$
Variance	S_n^2	$\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X})^2$

Point Estimation VI

- Method of maximum likelihood:
  It consists in choosing as estimator of $\underline{\theta}$ that statistic, which when substituted for $\underline{\theta}$, maximizes the likelihood function

$$L = f_{\underline{X}}(\underline{x}, \underline{\theta}) = f_{X_1,\ldots,X_n}(x_1, \ldots, x_n, \theta_1, \ldots, \theta_l).$$

- Procedure to find maximum likelihood estimate (mle):
  - Define the likelihood function, $L(\theta)$.
  - Often it is easier to take the natural logarithm (ln) of $L(\theta)$.
  - When applicable, differentiate $l(\theta) = lnL(\theta)$ with respect to θ, and then equate the derivative to zero.
  - Solve for the parameter θ, and we will obtain $\hat{\theta}$.
  - Check whether it is a maximizer or global maximizer.

Point Estimation VII

- MLE for Bernoulli, Exponential, Normal
- Invariance property of maximum likelihood estimators
  Let $h(\theta)$ be a one-to-one function of θ. If $\underline{\hat{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_l)$ is the MLE of $\theta = (\theta_1, \ldots, \theta_l)$, then the MLE of a function

  $$h(\underline{\theta}) = (h(\theta_1), \ldots, h(\theta_l))$$

  of these parameters is $h(\underline{\hat{\theta}}) = (h(\hat{\theta}_1), \ldots, h(\hat{\theta}_l))$ for $1 \leq k \leq l$.
    - MLE of σ is $\sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$ since it was shown that the MLE of σ^2 is $\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\right)$.

Point Estimation VIII

Some Desirable Properties of Point Estimators

- Unbiased
- Sufficiency
- Consistency
- Efficiency

Unbiased Estimators

- A point estimator $\hat{\theta}$ is called an unbiased estimator of the parameter θ if

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

  for all possible values of θ.
- Otherwise $\hat{\theta}$ is said to be biased.
  - Furthermore, the bias of $\hat{\theta}$ is given by

$$B(\hat{\theta}) = E[\hat{\theta}] - \theta.$$

Unbiased Estimators II

- *Theorems*:
  - The sample mean, $\bar{x}\left(=\frac{1}{n}\sum_{i=1}^{n}X_i\right)$ is an unbiased estimator of the population mean μ.
    - Sketch of proof: $E[\bar{X}]=E\left(\frac{1}{n}\sum_{i=1}^{n}X_i\right)=\mu$

Unbiased Estimators III

- The statistic, $S_n^2 \left(= \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)$ is not an unbiased estimator of the population variance σ^2.
  - Sketch of proof:

$$
\begin{aligned}
E\left[S_n^2\right] &= E\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2\right) = \frac{1}{n}E\left(\sum_{i=1}^{n}\left((X_i - \mu) - (\bar{X} - \mu)\right)^2\right) \\
&= \frac{1}{n}\left(\sum_{i=1}^{n} E(X_i - \mu)^2 - nE(\bar{X} - \mu)^2\right) = \frac{1}{n}\left(n\sigma^2 - n\frac{\sigma^2}{n}\right) = \left(\frac{n-1}{n}\right)\sigma^2
\end{aligned}
$$

- The sample variance, $S^2 \left(= \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2 \right)$ is an unbiased estimator of the population variance σ^2.
  - Sketch of proof: $E\left[S^2\right] = E\left[\frac{n}{n-1} S_n^2\right] = \frac{n}{n-1} E\left[S_n^2\right] = \sigma^2$

Unbiased Estimators IV

- Unbiased estimators need not be unique.
  - Let $X_1, \ldots, X_1 n$ be a random sample from a population with finite mean μ. Then the sample mean $\bar{X}$ and $\frac{1}{3}\bar{X} + \frac{2}{3}X_1$ are both unbiased estimators of μ.
  - Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators of θ. Then $\hat{\theta}_3 = a\hat{\theta}_1 + (1 - a)\hat{\theta}_2, 0 \leq a \leq 1$ is an unbiased estimator of θ.

Mean Square Error of an Estimator

- The mean square error of the estimator $\hat{\theta}$, denoted by $MSE(\hat{\theta})$, is defined as

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

- *Theorem*: $MSE(\hat{\theta}) = Var(\hat{\theta}) + B^2(\hat{\theta})$.
  - Sketch of proof:

$$\begin{aligned} MSE(\hat{\theta}) &= E(\hat{\theta} - \theta)^2 = E\left((\hat{\theta} - E(\hat{\theta})) + (E(\hat{\theta}) - \theta)\right)^2 \\ &= E\left(\hat{\theta} - E(\hat{\theta})\right)^2 + \left(E(\hat{\theta}) - \theta\right)^2 + 2\left(E(\hat{\theta}) - \theta\right)\left(E(\hat{\theta}) - E(\hat{\theta})\right) \\ &= Var(\hat{\theta}) + B^2(\hat{\theta}). \end{aligned}$$

Unbiased Estimators VI

- If $\hat{\theta}$ is an unbiased estimator of θ, then

$$B^2(\hat{\theta}) = 0 \text{ and } MSE(\hat{\theta}) = Var(\hat{\theta}).$$

- Minimum variance unbiased estimator (MVUE) of θ :
  - The unbiased estimator $\hat{\theta}$ that minimizes the mean square error is called the *MVUE* of θ.

Sufficient Estimators

- A statistic U is a sufficient statistic for a parameter θ if U contains all the information available in the data about the value of θ.
- If $U(\mathbf{X})$ is a sufficient statistic for a parameter θ, then any inference about θ should depend on the sample $\mathbf{X}$ only through the value $U(\mathbf{X})$.
    - That is, if $\mathbf{x}$ and $\mathbf{y}$ are two sample points such that $U(\mathbf{x}) = U(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{X} = \mathbf{y}$ is observed.
- Example
    - the sample mean ($\bar{X}$) may contain all the relevant information about the parameter μ, and in that case $U(\mathbf{X}) = \bar{X}$ is called a sufficient statistic for μ.

Formal definition

- Let $X_1, \ldots, X_n$ be a random sample from a probability distribution with unknown parameter θ.
  - Then, the statistic $U = g(X_1, \ldots, X_n)$ is said to be **sufficient statistic** for θ if $f_{X_1,\ldots,X_n}(x_1, \ldots, x_n | U = u)$ does not depend on θ for any value of u.
  - An estimator of θ that is a function of a sufficient statistic for θ is said to be a **sufficient estimator** of θ.

- Example: Let $X_1, \ldots, X_n$ be i.i.d. Bernoulli random variables with parameter θ. Then $U = \sum_{i=1}^{n} X_i$ is sufficient for θ.
- Sketch of proof:

$$\begin{aligned} f(X_1, \ldots, X_n; \theta) &= \theta^{\sum_{i=1}^{n} X_i} (1-\theta)^{n - \sum_{i=1}^{n} X_i} \\ &= \theta^U (1-\theta)^{n-U}. \end{aligned}$$

Since, $U \sim Bin(n, \theta)$,

$$f(U; \theta) = {}^nC_u \theta^U (1-\theta)^{n-U}.$$

Thus,

$$f(x_1, \ldots, x_n | U = u) = \frac{f(x_1, \ldots, x_n, u)}{f_U(u)} = \begin{cases} \frac{1}{{}^nC_u} & ; \text{ if } u = \sum x_i \\ 0 & ; \text{ otherwise.} \end{cases}$$

Sufficient Estimators IV

Neyman–Fisher factorization theorem to *spot* a sufficient statistic.

- *Theorem*: -
  Let U be a statistic based on the random sample $X_1, \ldots, X_n$.
  Then, U is a sufficient statistic for θ if and only if the joint p.d.f or p.m.f., $f(x_1, , \ldots, x_n; \theta)$ can be factored into two non-negative functions, i.e.,

$$f(x_1, \ldots, x_n; \theta) = g(u, \theta) h(x_1, \ldots, x_n), \text{ for all } x_1, \ldots, x_n,$$

  where
  - $g(u, \theta)$ is a function only of u and θ
  - and $h(x_1, \ldots, x_n)$ is a function of only $x_1, \ldots, x_n$ and not of θ.

Sufficient Estimators V

- Sketch of proof: discrete case
  - Sufficient => Factorization

$$
\begin{aligned}
f(x_1, \ldots, x_n; \theta) &= P_\theta(X_1 = x_1, \ldots, X_n = x_n, U = u) \\
&= P_\theta(X_1 = x_1, \ldots, X_n = x_n | U = u) P_\theta(U = u) \\
&\overset{suff.}{=} h(x_1, \ldots, x_n) g(u, \theta).
\end{aligned}
$$

  - Factorization => Sufficient

$$
\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n | U = u) &= \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, U = u)}{P_\theta(U = u)} \\
&= \begin{cases} \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, U = u)}{P_\theta(U = u)} & \text{if } (x_1, \ldots, x_n) \in A_u \\ 0 & \text{if } (x_1, \ldots, x_n) \notin A_u, \end{cases}
\end{aligned}
$$

where A_u is the set of all $(x_1, \ldots, x_n)$ such that U maps it into u.

Sufficient Estimators VI

- When $(x_1, x_2, \ldots, x_n) \in A_u$,

$$\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n | U = u) &= \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, U = u)}{P_\theta(U = u)} \\
&= \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n)}{P_\theta(U = u)} \\
&= \frac{f(x_1, \ldots, x_n, \theta)}{\sum\limits_{(x_1, \ldots, x_n) \in A_u} f(x_1, \ldots, x_n, \theta)} \\
&\overset{fact}{=} \frac{g(u, \theta) h(x_1, \ldots, x_n)}{\sum\limits_{(x_1, \ldots, x_n) \in A_u} g(u, \theta) h(x_1, \ldots, x_n)} \\
&= \frac{h(x_1, \ldots, x_n)}{\sum\limits_{(x_1, \ldots, x_n) \in A_u} h(x_1, \ldots, x_n)} \perp \theta
\end{aligned}$$

- When $(x_1, x_2, \ldots, x_n) \notin A_u$,

$$\begin{aligned}
P_\theta(X_1 = x_1, \ldots, X_n = x_n | U = u) &= \frac{P_\theta(X_1 = x_1, \ldots, X_n = x_n, U = u)}{P_\theta(U = u)} \\
&= 0 \perp \theta
\end{aligned}$$

Sufficient Estimators VII

- Procedure to verify Sufficiency
  1. Obtain the joint pdf or pmf $f_\theta(x_1, \ldots, x_n)$.
  2. If necessary, rewrite the joint pdf or pmf in terms of the given statistic and parameter so that one can use the factorization theorem.
  3. Define the functions g and h, in such a way that g is a function of the statistic and parameter only and h is a function of the observations only.
  4. If step 3 is possible, then the statistic is sufficient. Otherwise, it is not sufficient.

Sufficient Estimators VIII

- Example: Let $X_1, \ldots, X_n$ denote a random sample from a geometric population with parameter p. Show that $\bar{X}$ is sufficient for p.
    - Sketch of proof:

$$
\begin{aligned}
f(x_1, \ldots, x_n) &= \prod_{i=1}^{n} p(1-p)^{x_i - 1} \\
&= p^n (1-p)^{-n + \sum_{i=1}^{n} x_i} \\
&= p^n (1-p)^{-n + n\bar{x}} \\
&= g(\bar{x}, p) \underbrace{h(x_1, \ldots, x_n)}_{1}
\end{aligned}
$$

Sufficient Estimators IX

- *Joint Sufficiency*:
  Two statistics U_1 and U_2 are said to be jointly sufficient for the parameters θ_1 and θ_2 if the conditional distribution of $X_1, \ldots, X_n$ given U_1 and U_2 does not depend on θ_1 or θ_2.
  In general, the statistic $U = (U_1, \ldots, U_n)$ is jointly sufficient for $\theta = (\theta_1, \ldots, \theta_n)$ if the conditional distribution of $X_1, \ldots, X_n$ given U is free of θ.

- Factorization criteria for Joint Sufficiency
  Theorem: -
  The two statistics U_1 and U_2 are jointly sufficient for θ_1 and θ_2 if and only if the likelihood function can be factored into two non-negative functions,

$$f(x_1, \ldots, x_n; \theta_1, \theta_2) = g(u_1, u_2; \theta_1, \theta_2)h(x_1, \ldots, x_n)$$

  where $g(u_1, u_2; \theta_1, \theta_2)$ is only a function of $u_1, u_2; \theta_1$ and θ_2, and $h(x_1, \ldots, x_n)$ is free of θ_1 or θ_2

Sufficient Estimators XI

- Example: Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.
  1. If μ is unknown and $\sigma^2 = \sigma_0^2$ is known, then $U_1 = \bar{X}$ is a sufficient statistic for μ.

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma_0}\right)^2} = \underbrace{(2\pi\sigma_0^2)^{-n/2} e^{-\frac{1}{2\sigma_0^2}\left[\sum_{i=1}^{n} x_i^2\right]}}_{h(x_1,\ldots,x_n)} \times \underbrace{e^{-\frac{1}{2\sigma_0^2}\left[-2n\mu\bar{x}+n\mu^2\right]}}_{g(u_1,\theta_1)}$$

Sufficient Estimators XII

2. If $\mu = \mu_0$ is known and σ^2 is unknown, then $U_1 = \sum_{i=1}^{n}(X_i - \mu_0)^2$ is a sufficient statistic for σ^2.

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu_0}{\sigma}\right)^2} = \underbrace{(2\pi)^{-n/2}}_{h(x_1,\ldots,x_n)} \times \underbrace{\sigma^{-n} e^{-\frac{\sum_{i=1}^{n}(x_i-\mu_0)^2}{2\sigma^2}}}_{g(u_1,\theta_1)}$$

3. If μ and σ^2 are both unknown, then $U_1 = \sum_{i=1}^{n} X_i$ and $U_2 = \sum_{i=1}^{n} X_i^2$ are jointly sufficient for μ and σ^2.

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x_i-\mu}{\sigma}\right)^2} = \underbrace{(2\pi)^{-n/2}}_{h(x_1,\ldots,x_n)} \times \underbrace{\sigma^{-n} e^{-\frac{1}{2\sigma^2}\left[\sum_{i=1}^{n} x_i^2 - 2\mu \sum_{i=1}^{n} x_i + n\mu^2\right]}}_{g(u_1,\theta_1)}$$

Sufficient Estimators XIV

- *Theorem*: -
  If U is a sufficient statistic for θ, then the maximum likelihood estimator of θ, if unique, is a function of U.
  - Sketch of proof.

$$f(x_1, ..., x_n; \theta) \overset{suff}{=} g(u, \theta)h(x_1, \ldots, x_n).$$

Thus, the joint pdf/pmf depends on θ only through the statistic U. To maximize L we need to maximize $g(U, \theta)$.

Sufficient Estimators XV

- *Theorem*: -
  Let $X_1, \ldots, X_n$ be a random sample from a population with pdf or pmf of the exponential form

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} \exp[k(x)c(\theta) + S(x) + d(\theta)] & , x \in B \\ 0 & , x \notin B, \end{cases}$$

  where B does not depend on the parameter θ. The statistic $U = \sum_{i=1}^{n} k(X_i)$ is sufficient for θ.

- Distributions in exponential form
  - *Bernoulli*(p): $p^x(1-p)^{1-x} = \exp\left[x \ln(\frac{p}{1-p}) + \ln(1-p)\right]; x = 0, 1.$
  - *Poisson*(λ): $\frac{\lambda^x e^{-\lambda}}{x!} = \exp\left[x \ln(\lambda) - \ln(x!) - \lambda\right]; x = 1, 2, \ldots.$
  - *Normal*$(\mu, 1)$: $\frac{1}{\sqrt{2\pi}} e^{-(x-\mu)^2/2} = \exp\left[x\mu - \frac{x^2}{2} - \frac{\mu^2}{2} - \frac{\ln(2\pi)}{2}\right]; x \in \mathcal{R}.$

Sufficient Estimators XVI

- Sketch of proof.

$$\begin{aligned} f(x_1,\ldots,x_n;\theta) &= \exp\left[c(\theta)\sum_{i=1}^{n}k(x_i)+\sum_{i=1}^{n}S(x_i)+nd(\theta)\right] \\ &= \underbrace{\exp\left[c(\theta)\sum_{i=1}^{n}k(x_i)+nd(\theta)\right]}_{g(u,\theta)}\underbrace{\exp\left[\sum_{i=1}^{n}S(x_i)\right]}_{h(x_1,\ldots,x_n)}. \end{aligned}$$

Sufficient Estimators XVII

Some Observations on Sufficiency

1. All the statistics need not be sufficient, in addition
2. Any function of a sufficient statistic needs not to be sufficient, however
3. Any one-to-one function of a sufficient statistic is also sufficient statistic

Sufficient Estimators XVIII

- *RAO–BLACKWELL Theorem*: -
  Let $X_1, \ldots, X_n$ be a random sample with joint pmf or pdf $f(x_1, \ldots, x_n; \theta)$ and let $U = (U_1, \ldots, U_n)$ be jointly sufficient for $\theta = (\theta_1, \ldots, \theta_n)$. If T is any unbiased estimator of $k(\theta)$, and if $T^* = E(T|U)$, then:
  - T^* is an unbiased estimator of $k(\theta)$.
    Sketch of proof: - $ET^* = E(E(T|U)) = E(T) = k(\theta)$.

    Hence, T^* is an unbiased estimator of $k(\theta)$.
  - T^* is a function of U, and does not depend on θ.

    Sketch of proof: - Because U is sufficient for θ, the conditional distribution of any statistic (hence, for T), given U, does not depend on θ.
  - $Var(T^*) \leq Var(T)$ for every θ, and $Var(T^*) < Var(T)$ for some θ unless $T^* = T$ with probability 1.
    Sketch of proof: - $Var(T) = E(Var(T|U)) + Var(E(T|U)) = E(Var(T|U)) + Var(T^*)$.
    Because $Var(T|U) \geq 0$ for all u, it follows that $E(Var(T|U)) \geq 0$. Hence, $Var(T^*) \leq Var(T)$.

    Also $Var(T^*) = Var(T)$ *iff* $Var(T|U) = 0$ or T is a function of U, in which case $T^* = E(T|U) = T$.

Sufficient Estimators XIX

More Observations

- If one is searching for an unbiased estimator with minimal variance, it has to be restricted to functions of a sufficient statistics.
- If $k(\theta) = \theta$, and T is an unbiased estimator of θ, then $T^* = E(T|U)$ will typically give the MVUE of θ.

- Minimal sufficient statistic:
  A sufficient statistic $T(X)$ is called a minimal sufficient statistic if for any other sufficient statistic $T'(X)$,
  $$T(X) = g(T'(X)),$$
  i.e., $T(X)$ is a function of $T'(X)$.
- Intuitively, a minimal sufficient statistic most efficiently captures all possible information about the parameter θ.

- Let $f(\mathbf{x}|\theta)$ be the pmf/pdf of a sample $\mathbf{X}$. Suppose there exists a function $T(\mathbf{x})$ such that, for every two sample points $\mathbf{x}$ and $\mathbf{y}$, the ratio
$$\frac{f(\mathbf{x}|\theta)}{f(\mathbf{y}|\theta)}$$
is constant as a function of θ if and only if
$$T(\mathbf{x}) = T(\mathbf{y}).$$
Then $T(\mathbf{X})$ is a minimal sufficient statistic for θ.

Sufficient Estimators XXII

- Lehmann and Scheffe method to find a minimal sufficient statistic
  - Let $X_1, \dots, X_n$ be a random sample with pdf or pmf $f(x)$ that depends on a parameter θ.
  - Let $(x_1, \dots, x_n)$ and $(y_1, \dots, y_n)$ be two different sets of values of $(X_1, \dots, X_n)$.
    - Let $\frac{L(\theta; x_1, \dots, x_n)}{L(\theta; y_1, \dots, y_n)}$ be the ratio of the likelihoods evaluated at these two points.
  - Suppose it is possible to find a function $g(x_1, \dots, x_n)$ such that this ratio will be free of the unknown parameter θ if and only if $g(x_1, \dots, x_n) = g(y_1, \dots, y_n)$,
    - in other words

$$\frac{L(\theta; x_1, \dots, x_n)}{L(\theta; y_1, \dots, y_n)} \text{ is independent of } \theta \Leftrightarrow g(x_1, \dots, x_n) = g(y_1, \dots, y_n).$$

  - If such a function g can be found, then $g(X_1, \dots, X_n)$ is a minimal sufficient statistic for θ.

Sufficient Estimators XXIII

- Example 1: -
  Let $(X_1, \ldots, X_n)$ be a random sample from the *Bernoulli*(p), where p is unknown. Find a minimal sufficient statistic for p.
  - Solution: - The ratio of the likelihoods is

$$\frac{L(x_1, \ldots, x_n)}{L(y_1, \ldots, y_n)} = \frac{p(x_1, \ldots, x_n)}{p(y_1, \ldots, y_n)} = \frac{p^{\sum x_i}(1-p)^{n-\sum x_i}}{p^{\sum y_i}(1-p)^{n-\sum y_i}} = \left(\frac{p}{1-p}\right)^{\sum x_i - \sum y_i}$$

This ratio is to be independent of p, if and only if

$$g(x_1, \ldots, x_n) = \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i = g(y_1, \ldots, y_n).$$

Therefore, $g(X_1, \ldots, X_n) = \sum_{i=1}^{n} X_i$ is a minimal sufficient statistic for p.

- Example 2: -
  Let $(X_1, \ldots, X_n)$ be a random sample from the *Normal*(μ, σ^2), where μ and σ^2 are unknown. Show that $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).
  - Solution: - The ratio of the likelihoods is

$$\begin{aligned}\frac{L(x_1, \ldots, x_n)}{L(y_1, \ldots, y_n)} &= \frac{(2\pi\sigma^2)^{-n/2}\exp(-[n(\bar{x}-\mu)^2+(n-1)S_x^2]/(2\sigma^2))}{(2\pi\sigma^2)^{-n/2}\exp(-[n(\bar{y}-\mu)^2+(n-1)S_y^2]/(2\sigma^2))} \\ &= \exp\left(\frac{-n\left[\bar{x}^2-\bar{y}^2\right]+2n\mu\left[\bar{x}-\bar{y}\right]-(n-1)\left[S_x^2-S_y^2\right]}{2\sigma^2}\right)\end{aligned}$$

This ratio is to be independent of (μ, σ^2), if and only if

$$\bar{x}=\bar{y} \text{ and } S_x^2=S_y^2$$

Therefore, $(\bar{X}, S^2)$ is a minimal sufficient statistic for (μ, σ^2).

Sufficient Estimators XXV

- Note that
  - The dimension of a minimal sufficient statistic may not match the dimension of the parameter.
    - Example 3: -
      Let $(X_1, \ldots, X_n)$ be a random sample from the *Uniform*$(\theta, \theta + 1)$, where $-\infty < \theta < \infty$ is unknown. Find a minimal sufficient statistic for θ.
    - Solution: - The joint pdf of $\mathbf{X}$ is

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} 1 & , \theta < x_i < \theta + 1, i = 1, \ldots, n. \\ 0 & , \textit{otherwise}. \end{cases}$$

which can be rewritten as

$$f(x_1, \ldots, x_n; \theta) = \begin{cases} 1 & , \max_i x_i - 1 < \theta < \min_i x_i. \\ 0 & , \textit{otherwise}. \end{cases}$$

Thus, for two sample points $\mathbf{x}$ and $\mathbf{y}$, the numerator and denominator of the ratio $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ will be positive for the same values of θ if and only if $\min_i X_i = \min_i y_i$ and $\max_i X_i = \max_i y_i$. And, if the minima and maxima are equal, then the ratio is constant and, in fact, equals 1.

Thus, letting $X_{(1)} = \min_i X_i$ and $X_{(n)} = \max_i X_i$, we have that $T(X) = (X_{(l)}, X_{(n)})$ is a minimal sufficient statistic.

- A minimal sufficient statistic is not unique.
- Any one-to-one function of a minimal sufficient statistic is also a minimal sufficient statistic.
  - Thus $\left(R = X_{(n)} - X_{(1)}, M = \left(X_{(n)} + X_{(1)}\right)/2\right)$, is also a minimal sufficient statistic in Example 3

Sufficient Estimators XXVII

- Ancillary Statistic: - A statistic $T(\mathbf{X})$ whose distribution does not depend on the parameter θ is called an ancillary statistic for θ.

Sufficient Estimators XXVIII

- Example: -
  Let $(X_1, \ldots, X_n)$ be a random sample from the *Uniform*$(\theta, \theta + 1)$, where $-\infty < \theta < \infty$ is unknown. One can show $R = x_{(n)} - x_{(1)}$ is an ancillary statistic for θ.

- Solution: - The cdf of X_i is

$$F(x|\theta) = \begin{cases} 0 & , x \leq \theta, \\ x - \theta & , \theta < x < \theta + 1, \\ 1 & , \theta + 1 \leq x. \end{cases}$$

Thus the joint pdf of $\left(x_{(1)}, x_{(n)}\right)$ is

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

Making the transformation $R = X_{(n)} - X_{(1)}$ and $M = \left(X_{(n)} + X_{(1)}\right)/2$, which has the inverse transformation $X_{(1)} = (2M - R)/2$ and $X_{(n)} = (2M + R)/2$, with Jacobian equals 1.
Thus, the joint pdf of (R, M) is

$$h(r, m|\theta) = \begin{cases} n(n-1)r^{n-2} & , 0 < r < 1, \theta + r/2 < m < \theta + 1 - r/2. \\ 0 & , \textit{otherwise}. \end{cases}$$

Thus, the pdf for R is

$$h(r|\theta) = \int_{\theta + r/2}^{\theta + 1 - r/2} n(n-1)r^{n-2} dm = n(n-1)r^{n-2}(1-r);\ 0 < r < 1.$$

Thus, the distribution of R does not depend on θ, and R is ancillary.

Sufficient Estimators XXX

- Note that
  - Since the distribution of an ancillary statistic does not depend on θ, it might be suspected that a minimal sufficient statistic is unrelated to (or mathematically speaking, functionally independent of) an ancillary statistic.
  - However, this is not necessarily the case. As we have seen in the last two examples that *R* is ancillary as well as part of minimal sufficient statistic.
  - Hence, the ancillary statistic and the minimal sufficient statistic are not independent.
  - For many important situations, however, the intuition that a minimal sufficient statistic is independent of any ancillary statistic is correct.
    - A minimal sufficient statistic is independent of any ancillary statistic if it a complete statistic as well.

- Complete Statistic: -
  Let $f(t|\theta)$ be a family of pdfs/pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called complete if $E_\theta g(T) = 0$ for all θ implies $P_\theta(g(T) = 0) = 1$ for all θ.
  Equivalently, $T(\mathbf{X})$ is called a complete statistic.

Sufficient Estimators XXXII

- Note that
  - Completeness is a property of a family of probability distributions, not of a particular distribution.
  - If $X \sim (0, 1)$, then defining $g(x) = x$, we have $Eg(X) = 0$. But the function $g(x) = x$ satisfies $P(g(X) = 0) = 0$, not 1.
  - However, this is a particular distribution, not a family of distributions.
  - If $X \sim N(\theta, 1), -\infty < \theta < \infty$, we can see that no function of X, except one that is 0 with probability 1 for all θ, satisfies $E_\theta g(X) = 0$ for all θ.
    - Thus, the family of $N(\theta, 1)$ distributions, $-\infty < \theta < \infty$, is complete.

Sufficient Estimators XXXIII

- Binomial complete sufficient statistic: -
  Suppose $T \sim Binomial(n, p)$,, with $0 < p < 1$. Let g be a function such that $E_p g(T) = 0$. Then,

$$\begin{aligned} 0 &= E_p g(T) = \sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t} \\ &= (1-p)^n \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t \end{aligned}$$

for all $p, 0 < p < 1$. Thus it must be that

$$0 = \sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t \Leftrightarrow g(t) = 0, \text{ for all } t = 0, \ldots, n.$$

Since, T takes values $0, 1, \ldots, n$ with probability 1, this yields that $P_p(g(T) = 0) = 1$, for all p. Hence, T is a complete statistic.

Sufficient Estimators XXXIV

- Basu's Theorem
  If $T(X)$ is a complete and minimal sufficient statistic, then $T(X)$ is independent of every ancillary statistic.
- Note that
  - To show that a statistic is complete, which is sometimes a rather difficult analysis problem.
  - Fortunately, most problems we are concerned with are covered by the following theorem.

Sufficient Estimators XXXV

- Complete statistics in the exponential family: -
  Let $X_1, \ldots, X_n$ be iid observations from an exponential family with pdf or pmf of the form

$$f(\mathbf{x}|\underline{\theta}) = h(\mathbf{x})c(\underline{\theta}) \exp\left(\sum_{j=1}^{k} w(\theta_j)t_j(\mathbf{x})\right),$$

  where $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ then the statistic

$$T(\mathbf{X}) = \left(\sum_{i=1}^{n} t_1(X_i), \sum_{i=1}^{n} t_2(X_i), \ldots \sum_{i=1}^{n} t_k(X_i)\right)$$

  is complete as long as the parameter space Θ contains an open set in $\mathcal{R}^k$.

Sufficient Estimators XXXVI

- Application of Basu's Theorem
  - Independence of $\bar{X}$ and S, the sample mean and variance, when sampling from a $N(\mu, \sigma^2)$ population.
    - For known σ^2, and unknown $\mu \in (-\infty, \infty)$, $\bar{X}$ is sufficient statistic for μ.
    - In addition, the family of $N(\mu, \sigma^2)$ distributions, where $-\infty < \mu < \infty$, and with known $\frac{\sigma^2}{n}$, is a complete family. Since this is the distribution of $\bar{X}$, $\bar{X}$ is a complete statistic.
    - The S is ancillary statistic for μ, since $(n-1)\dfrac{S^2}{\sigma^2} \sim \chi^2_{n-1}$.

Sufficient Estimators XXXVII

- Note
  - If a minimal sufficient statistic exists, then any complete statistic is also a minimal sufficient statistic.
  - So even though the word “minimal" is redundant in the statement of Basu’s Theorem, it was stated in this way as a reminder that the statistic $T(X)$ in the theorem is a minimal sufficient statistic.

Consistent Estimators

- A statistic is a consistent estimator if its value becomes closer to the value of the true parameter which is being estimated, as the sample size becomes larger.

Consistent Estimators II

Formal definition

- The estimator $\hat{\theta}_n$ is said to be a consistent estimator of θ if, for any $\epsilon > 0$,

$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| \leq \epsilon\right] = 1$$

or equivalently,

$$\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| > \epsilon\right] = 0,$$

i.e. $\hat{\theta}_n$ converges in probability to θ.

Consistent Estimators III

- Example: - Let $X_1, \ldots, X_n$ be a random sample with true mean μ and finite variance, σ^2. Then, the sample mean $\bar{X}$ is a consistent estimator of the population mean μ.
  - Sketch of proof: -
    Note that, for any positive r.v. X (Markov Inequality)

$$E[X] = \int_0^{\epsilon} x f_X(x)dx + \int_{\epsilon}^{\infty} x f_X(x)dx \geq \int_0^{\epsilon} x f_X(x)dx + \epsilon \int_{\epsilon}^{\infty} f_X(x)dx \geq \epsilon P(X \geq \epsilon).$$

Hence, (Chebyshev's Inequality)

$$P\left[|\bar{X} - \mu| \geq \epsilon\right] = P\left[(\bar{X} - E[\bar{X}])^2 \geq \epsilon^2\right] \leq \frac{E(\bar{X} - E[\bar{X}])^2}{\epsilon^2} = \frac{Var(\bar{X})}{\epsilon^2},$$

Thus,

$$\lim_{n\to\infty} P\left[|\bar{X} - \mu| \geq \epsilon\right] \leq \lim_{n\to\infty} \frac{Var(\bar{X})}{\epsilon^2} = \lim_{n\to\infty} \frac{\sigma^2}{n\epsilon^2} = 0$$

Test for Consistency (Sufficient conditions)

- *Theorem*: - An unbiased estimator $\hat{\theta}_n$ of θ is a consistent estimator for θ if

$$\lim_{n\to\infty} Var(\hat{\theta}_n) = 0.$$

- Sketch of proof: - $\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| \geq \epsilon\right] \overset{U.E.}{=} \lim_{n\to\infty} P\left[|\hat{\theta}_n - E(\hat{\theta}_n)| \geq \epsilon\right] \overset{C.I.}{\leq} \lim_{n\to\infty} \frac{Var(\hat{\theta}_n)}{\epsilon^2} = 0$
- An unbiased estimator $\hat{\theta}_n$ is consistent if $Var(\hat{\theta}_n) \to 0$ as $n \to \infty$.

Consistent Estimators V

(*Contd.*)

- *Theorem*: - An estimator $\hat{\theta}_n$, with finite variance, is a consistent estimator for θ if
$$\lim_{n\to\infty} E(\hat{\theta}_n - \theta)^2 = 0.$$
    - Sketch of proof: - $\lim_{n\to\infty} P\left[|\hat{\theta}_n - \theta| \geq \epsilon\right] \overset{C.I.}{\leq} \lim_{n\to\infty} \frac{E(\hat{\theta}_n - \theta)^2}{\epsilon^2} = 0$
    - A biased estimator $\hat{\theta}_n$ is consistent if both $Var(\hat{\theta}_n) \to 0$ and $B(\hat{\theta}_n) \to 0$ as $n \to \infty$.

Consistent Estimators VI

- Example: -
  Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$ population.
  - Then the sample variance S^2 is a consistent estimator for σ^2.
    - Sketch of proof: -
      $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2(n-1) \Rightarrow Var\left[(n-1)\frac{S^2}{\sigma^2}\right] = 2(n-1) \Rightarrow Var(S^2) = \frac{2\sigma^4}{n-1}$.
      S^2 is u.e. of σ^2 and $\lim_{n\to\infty} Var(S^2) = 0$.
  - The maximum likelihood estimators $\bar{X}$ and S_n^2 for μ and σ^2, respectively, are consistent estimators for μ and σ^2.
    - Sketch of proof: -
      $\bar{X}$ is u.e. of μ and $\lim_{n\to\infty} Var(\bar{X}) = 0$, thus $\bar{X}$ is consistent estimator of μ.
      $Var(S_n^2) = Var\left(\frac{n-1}{n} S^2\right) = \frac{2(n-1)\sigma^4}{n^2}$, thus $\lim_{n\to\infty} Var(S_n^2) = 0$ and
      $B(S_n^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2$, thus $\lim_{n\to\infty} B(S_n^2) = 0$

Efficiency I

Efficiency

- It is a relative comparison between variances of two unbiased/biased estimators

Efficiency II

Formal Definitions

- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two unbiased estimators for θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is the ratio

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{Var(\hat{\theta}_2)}{Var(\hat{\theta}_1)}.$$

  - If $Var(\hat{\theta}_1) < Var(\hat{\theta}_2)$, or equivalently, $e(\hat{\theta}_1, \hat{\theta}_2) > 1$, then, $\hat{\theta}_1$ is relatively more efficient than $\hat{\theta}_2$.

Efficiency III

(*Contd.*)

- If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two biased estimators for θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is the ratio

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{E(\hat{\theta}_2 - \theta)^2}{E(\hat{\theta}_1 - \theta)^2} = \frac{MSE(\hat{\theta}_2)}{MSE(\hat{\theta}_1)}.$$

  - If $MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2)$, or equivalently, $e(\hat{\theta}_1, \hat{\theta}_2) > 1$, then, $\hat{\theta}_1$ is relatively more efficient than $\hat{\theta}_2$.

Efficiency IV

- Example:-
  Let $X_1, \ldots, X_n, n \geq 2$ be a random sample from a normal population with a true mean μ and variance σ^2. Consider the following two estimators of σ^2: $\theta_1 = S^2$, and $\theta_2 = S_n^2$. Find $e(\theta_1, \theta_2)$.
  - Sketch of proof: -

$$
\begin{aligned}
MSE(\hat{\theta}_1) &= 0 + \frac{2\sigma^4}{n-1} = \frac{2\sigma^4}{n-1} \\
MSE(\hat{\theta}_2) &= \left(-\frac{\sigma^2}{n}\right)^2 + \frac{2(n-1)\sigma^4}{n^2} = \frac{(2n-1)\sigma^4}{n^2} \\
e(\hat{\theta}_1, \hat{\theta}_2) &= \frac{(n-1)(2n-1)}{2n^2} < 1, \text{ for } n \geq 2.
\end{aligned}
$$

Hence, S_n^2 is relatively more efficient than S^2.

Uniformly Minimum Variance Unbiased Estimator

- An unbiased estimator $\hat{\theta}_0$, is said to be a uniformly minimum variance unbiased estimator (UMVUE) for the parameter θ if, for any other unbiased estimator $\hat{\theta}$,

$$\mathit{Var}(\hat{\theta}_0) \leq \mathit{Var}(\hat{\theta}),$$

  for all possible values of θ.

Efficiency VI

Cramer–Rao Inequality:
(Lower bound for the variance of any estimator)

- *Theorem*: - Let $X_1, \ldots, X_n$ be a sample with pdf $f(\mathbf{x}|\theta)$, and let $W(\mathbf{X}) = W(X_1, \ldots, X_n)$ be any estimator satisfying

  $$\frac{d}{d\theta} E_\theta[W(\mathbf{X})] = \int_{\mathcal{X}} \frac{\delta}{\delta\theta} [W(\mathbf{x}) f(\mathbf{x}|\theta)] d\mathbf{x}$$

  and

  $$Var_\theta[W(\mathbf{X})] < \infty,$$

  then

  $$Var_\theta[W(\mathbf{X})] \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(\mathbf{X})]\right)^2}{E_\theta\left[\left(\frac{\delta}{\delta\theta} \ln f(\mathbf{x}|\theta)\right)^2\right]}.$$

Efficiency VII

- Sketch of proof: -
  Note that

$$
\begin{aligned}
\frac{d}{d\theta}E_\theta[W(\mathbf{X})] &= \int_{\mathcal{X}} \frac{\delta}{\delta\theta}[W(\mathbf{x})f(\mathbf{x}|\theta)]d\mathbf{x} \\
&= \int_{\mathcal{X}} W(\mathbf{x})\left[\frac{\delta}{\delta\theta}f(\mathbf{x}|\theta)\right] d\mathbf{x} \\
&= \int_{\mathcal{X}} W(\mathbf{x})\left[\frac{\delta f(\mathbf{x}|\theta)}{\delta\theta}\frac{1}{f(\mathbf{x}|\theta)}\right] f(\mathbf{x}|\theta)d\mathbf{x} \\
&= E_\theta\left[W(\mathbf{X})\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right] \\
&= Cov_\theta\left(W(\mathbf{X}), \frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right),
\end{aligned}
$$

since $$E_\theta\left[\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right] = \int_{\mathcal{X}}\left[\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right] f(\mathbf{x}|\theta)d\mathbf{x} = \int_{\mathcal{X}}\left[\frac{\delta}{\delta\theta}f(\mathbf{x}|\theta)\right] d\mathbf{x} = \frac{d}{d\theta}\left[\int_{\mathcal{X}} f(\mathbf{x}|\theta)d\mathbf{x}\right] = 0.$$

Also

$$Var_\theta\left(\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right) = E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right)^2\right].$$

Now,

$$\frac{\left[Cov_\theta\left(W(\mathbf{X}), \frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right)\right]^2}{Var_\theta[W(\mathbf{x})]Var_\theta\left[\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right]} \leq 1 \Rightarrow \frac{\left[\frac{d}{d\theta}E_\theta[W(\mathbf{X})]\right]^2}{E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right)^2\right]} \leq Var_\theta\left[W(\mathbf{x})\right]$$

Efficiency VIII

- *Corollary*: -
  Let $X_1, \ldots, X_n$ be an iid random sample from a population with pdf or pmf $f_\theta(x)$ that depends on a parameter θ. If $\hat{\theta} = W(\mathbf{x})$ is an unbiased estimator of $\psi(\theta)$, then

$$Var(\hat{\theta}) \geq \frac{\left[\frac{d}{d\theta}\psi(\theta)\right]^2}{nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]}.$$

Efficiency IX

- Sketch of proof: - Note that

$$
\begin{aligned}
E\left[\left(\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right)^2\right] &= E\left[\left(\frac{\delta}{\delta\theta}\ln\prod_{i=1}^{n} f_\theta(x_i)\right)^2\right] = E\left[\left(\sum_{i=1}^{n}\frac{\delta}{\delta\theta}\ln f_\theta(x_i)\right)^2\right] \\
&= \sum_{i=1}^{n} E\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x_i)\right)^2\right] + \sum_{i\neq j} E\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x_i)\right)\left(\frac{\delta}{\delta\theta}\ln f_\theta(x_j)\right)\right] \\
&= nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right] + \sum_{i\neq j} E\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x_i)\right)\right] E\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x_j)\right)\right] \\
&= nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]
\end{aligned}
$$

and

$$
\left[\frac{d}{d\theta}E_\theta[W(\mathbf{x})]\right]^2 = \left[\frac{d}{d\theta}E_\theta[\hat{\theta}]\right]^2 = \left[\frac{d}{d\theta}\psi(\theta)\right]^2.
$$

Hence,

$$
Var(\hat{\theta}) \geq \frac{\left[\frac{d}{d\theta}E_\theta[\hat{\theta}]\right]^2}{E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f(\mathbf{x}|\theta)\right)^2\right]} = \frac{\left[\frac{d}{d\theta}\psi(\theta)\right]^2}{nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]}.
$$

- *Corollary*: -
  Let $X_1, \ldots, X_n$ be an iid random sample from a population with pdf or pmf $f_\theta(x)$ that depends on a parameter θ. If $\hat{\theta}$ is an unbiased estimator of θ, then

$$Var(\hat{\theta}) \geq \frac{1}{nE\left[\left(\frac{\delta}{\delta\theta} \ln f_\theta(x)\right)^2\right]}.$$

Efficient Estimator

- If $\hat{\theta}$ is an unbiased estimator of θ and if

$$Var(\hat{\theta}) = \frac{1}{nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]}$$

then $\hat{\theta}$ is a uniformly minimum variance unbiased estimator (UMVUE) of θ.
Sometimes $\hat{\theta}$ is also referred to as an efficient estimator.

Efficiency XII

- *Result*:- If the function $f(\cdot)$ is sufficiently smooth, specifically if $\frac{d}{d\theta}E_\theta\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right) = \int \frac{\delta}{\delta\theta}\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right) f_\theta(x)\right] dx$, then

$$E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right] = -E_\theta\left(\frac{\delta^2}{\delta\theta^2}\ln f_\theta(x)\right) = Var\left[\frac{\delta}{\delta\theta}\ln f_\theta(x)\right]$$

and for an unbiased estimator $\hat{\theta}$ for θ the Cramer-Rao inequality can be rewritten as

$$Var(\hat{\theta}) \geq \frac{1}{-nE\left(\frac{\delta^2}{\delta\theta^2}\ln f_\theta(x)\right)} = \frac{1}{nVar\left[\frac{\delta}{\delta\theta}\ln f_\theta(x)\right]}.$$

- Sketch of proof: -

$$
\begin{aligned}
E_\theta\left(\frac{\delta^2}{\delta\theta^2}\ln f_\theta(x)\right) &= E_\theta\left[\frac{\delta}{\delta\theta}\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)\right] = E_\theta\left[\frac{\delta}{\delta\theta}\left(\frac{\frac{\delta}{\delta\theta}f_\theta(x)}{f_\theta(x)}\right)\right] \\
&= E_\theta\left[\left(\frac{\frac{\delta^2}{\delta\theta^2}f_\theta(x)}{f_\theta(x)}\right) - \left(\frac{\frac{\delta}{\delta\theta}f_\theta(x)}{f_\theta(x)}\right)^2\right] \\
&= \int \frac{\delta^2}{\delta\theta^2}f_\theta(x)dx - E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right] \\
&\overset{assumption}{=} \frac{\delta}{\delta\theta}\int \frac{\delta}{\delta\theta}f_\theta(x)dx - E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right] \\
&= \frac{d}{d\theta}E_\theta\left[\frac{\delta}{\delta\theta}\ln f_\theta(x)\right] - E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right] \\
&= -E_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]
\end{aligned}
$$

Hence,

$$
Var(\hat\theta) \geq \frac{1}{nE_\theta\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]} = \frac{1}{-nE_\theta\left(\frac{\delta^2}{\delta\theta^2}\ln f_\theta(x)\right)} = \frac{1}{nVar_\theta\left[\frac{\delta}{\delta\theta}\ln f_\theta(x)\right]}.
$$

- Example: - Let $X_1, \ldots, X_n$ be a random sample from an $N(\mu, \sigma^2)$ population. Then $\hat{X}$ is an efficient estimator for μ.
  - Sketch of proof: -

$$l(x, \mu) = \ln f(x, \mu) = c - \frac{(x-\mu)^2}{2\sigma^2}.$$

Thus,

$$\frac{\delta}{\delta\mu} l(x, \mu) = \frac{x-\mu}{\sigma^2} \text{ and } \frac{\delta^2}{\delta\mu^2} l(x, \mu) = -\frac{1}{\sigma^2}.$$

Hence,

$$\frac{1}{nE\left[\left(\frac{\delta}{\delta\theta}\ln f_\theta(x)\right)^2\right]} = \frac{1}{nE\left[\left(\frac{\delta}{\delta\mu} l(x, \mu)\right)^2\right]} = \frac{1}{-nE\left[\frac{\delta^2}{\delta^2\mu} l(x, \mu)\right]} = \frac{\sigma^2}{n} = Var(\bar{X})$$

Efficiency XV

Note: -

- For a given problem UMVUE may not exist.
- Even when an UMVUE exists, it is not necessary that it have a variance equal to the Cramer–Rao lower bound.
- The term $I(\theta) = E\left[\left(\frac{\delta}{\delta\theta} \ln f_\theta(x)\right)^2\right]$ is called the Fisher information.
- It can be shown that the Fisher information in a sample of size n, denoted by $I_n(\theta)$, is n times the Fisher information in one observation. That is,
$$I_n(\theta) = nI(\theta).$$