7.3.2 Best unBiased Estimator's

if we find an unbiased Estimator with uniformly smallest variance - a best unbiased estimator - then our task is done.

The results are actually more general

Let W^* is an estimator of θ with

$$E_\theta[W^*] = \tau(\theta) \neq \theta$$, and we are interested in investigating the worth of W^*

$$C_\tau = \{W : E_\theta[W] = \tau(\theta)\}$$

Class of Estimator's.

For any $W_1, W_2 \in C_\tau$, $\text{Bias}_\theta W_1 = \text{Bias}_\theta W_2$

so $E_\theta[(W_1-\theta)^2] - E_\theta[(W_2-\theta)^2] = \text{Var}\, W_1 - \text{Var}\, W_2$

and MSE comparison's, within the class C_τ, can be based on variance alone.

Thus, although we speak about unbiased estimator's, we really are comparing estimator's that have the same expected value, $\tau(\theta)$.

Definition 7.3.7:-

An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $\mathbb{E}_\theta[W^*] = \tau(\theta)$ $\forall \theta$ and, for any other estimator W with $E_\theta(W) = \tau(\theta)$, we have $\mathrm{Var}_\theta W^* \le \mathrm{Var}_\theta W$ for all θ. W^* is also called a uniform minimum variance unbiased estimator (UMVUE) of $\tau(\theta)$.

Finding a best unbiased estimator (if one exists) is not an easy task for a variety of reason's.

Ex: 7.3.8 (Poisson unbaised estimation)

$X_1, X_2, \dots, X_n \sim \text{Poisson}(\lambda)$

$E_\lambda[\bar{X}] = \lambda \quad \forall \lambda, \quad E_\lambda[S^2] = \lambda \quad \forall \lambda$

But $\quad \text{var}_\lambda(\bar{X}) \leq \text{var}_\lambda[S^2]$

Consider a class of Estimator's

$$W_a(\bar{X}, S^2) = a\bar{X} + (1-a)S^2$$

$$\mathbb{E}_\lambda[W_a(\bar{X}, S^2)] = a\,E_\lambda[\bar{X}] + (1-a)\,\mathbb{E}_\lambda[S^2]$$

$$= a\lambda + (1-a)\lambda$$

$$= \lambda \quad \text{(unbaised Estimator)}$$

$\forall a \quad E[W_a(\bar{X}, S^2)] = \lambda$

$\Rightarrow$ we have infinitely many unbaised Estimator's for λ.

Eeven if $\bar{x}$ is better than S^2, is it Better than every $W_a(\bar{x}, S^2)$?

Futhermore How can we be sure that there are not other, better, unbaised Estimators lurking about?.

Suppose that, for Estimating a Parameter $\tau(\theta)$ of a distribution $f(x|\theta)$, we Can Specify a lower Bound, say $B(\theta)$ on the variance of any unbaised estimator of $\tau(\theta)$. if we can find an Estimator W^* satisfying $Var_\theta(W^*) = B(\theta)$, we have found a Best unbaised Estimator.

Score function is $\frac{\partial}{\partial\theta} l(\theta|x)$

$\Rightarrow \frac{\partial}{\partial\theta} \log L(\theta|x)$

First identity :- $\mathbb{E}\left[\frac{\partial l}{\partial\theta}\right] = 0$

2nd identity :- $\mathbb{E}\left[\frac{\partial^2 l}{\partial\theta^2}\right] + \mathbb{E}\left[\left(\frac{\partial l}{\partial\theta}\right)^2\right] = 0$

Obtained from $\int f_\theta(y)\,dy \equiv 1$

Proof:

$$\frac{\partial}{\partial\theta} \log f(x|\theta) = \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)}$$

$$\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right] = \int_x \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{\cancel{f(x|\theta)}} \cancel{f}(x|\theta)\,dx$$

$$= \int_x \frac{\partial}{\partial\theta} f(x|\theta)\,dx$$

$$= \frac{\partial}{\partial\theta} \int_x f(x|\theta)\,dx = \frac{\partial}{\partial\theta} 1 = 0$$

$$\Rightarrow \quad \mathbb{E}\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right] = 0$$

2nd identity:

$$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = \frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f(x|\theta)}{f(x|\theta)}\right]$$

$$= \frac{f(x|\theta)\frac{\partial^2}{\partial\theta^2}f(x|\theta) - \left(\frac{\partial}{\partial\theta}f(x|\theta)\right)^2}{f(x|\theta)^2}$$

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] = \underbrace{\int_x \frac{\partial^2}{\partial\theta^2}f(x|\theta)\,dx}_{\Rightarrow 0} - \int_x \frac{\left(\frac{\partial}{\partial\theta}f(x|\theta)\right)^2}{f(x|\theta)^2} f(x|\theta)\,dx$$

$$\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] = -\int_x \left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2 f(x|\theta)\,dx$$

$$\Rightarrow \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] = -\mathbb{E}\left[\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]^2\right]$$

$$\Rightarrow \mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right] + \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right] = 0$$

Theorem 7.3.9 (CRAMER- RAO Inequality)

Let $X_1, X_2, \ldots, X_n$ be a sample with pdf $f(x|\theta)$, and let $W(X) = W(X_1, \ldots X_n)$ be any Estimator satisfying

$$\frac{d}{d\theta} E_\theta [W(X)] = \int_x \frac{\partial}{\partial\theta} [W(x) f(x|\theta)] \, dx$$

and $\quad \mathrm{var}_\theta W(X) < \infty$

then

$$\mathrm{var}_\theta (W(X)) \geq \frac{\left(\frac{d}{d\theta} E_\theta (W(X))\right)^2}{E_\theta \left[\left(\frac{\partial}{\partial\theta} \log f(X|\theta)\right)^2 \right]}$$

Proof:

$$\mathrm{cov}(X, Y)^2 \leq \mathrm{Var}(X) \, \mathrm{var}(Y)$$

take $X = W(x)$, $Y = \frac{\partial}{\partial\theta} \log f(x|\theta)$

in above eq^n

we know $\mathbb{E}[Y]=0$ $\mathbb{E}[Y^2] + \mathbb{E}\left[\frac{\partial}{\partial\theta} Y\right] = 0$

$$\frac{d}{d\theta}\mathbb{E}[W(x)] = \int_x W(x)\left[\frac{\partial}{\partial\theta} f(x|\theta)\right] dx$$

$$= \int_x W(x) \frac{\frac{\partial}{\partial\theta} f(x|\theta)}{f(x|\theta)} f(x|\theta)\, dx$$

$$= \int_x W(x) \frac{\partial}{\partial\theta} \log f(x|\theta)\, f(x|\theta)\, dx$$

$$= \mathbb{E}\left[W(x) \frac{\partial}{\partial\theta} \log f(x|\theta)\right]$$

& $\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f(x|\theta)\right] = 0$

$$\Rightarrow \operatorname{cov}\left(W(x), \frac{\partial}{\partial\theta} \log f(x|\theta)\right) = \mathbb{E}\left[W(x) \frac{\partial}{\partial\theta} \log f(x|\theta)\right] - \mathbb{E}[W(x)]\, \underbrace{\mathbb{E}\left[\frac{\partial}{\partial\theta} \log f(x)\right]}_{0}$$

$$\Rightarrow \operatorname{cov}\left(W(x), \frac{\partial}{\partial\theta}\log f(x|\theta)\right)$$

$$= \mathbb{E}\left[W(x)\frac{\partial}{\partial\theta}\log f(x|\theta)\right]$$

$$= \frac{\partial}{\partial\theta}\mathbb{E}_\theta[W(x)]$$

take $X = W(x)$ $\quad Y = \frac{\partial}{\partial\theta}\log f(x|\theta)$ in the

$$\operatorname{cov}(x,y)^2 \leq \operatorname{var}(x)\operatorname{var}(y)$$

$$\Rightarrow \quad \operatorname{var}(x) \geq \frac{\operatorname{cov}(x,y)^2}{\operatorname{var}(y)}$$

$$\Rightarrow \quad \operatorname{var}(W(x)) \geq \frac{\left(\frac{\partial}{\partial\theta}\mathbb{E}[W(x)]\right)^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]}$$

$$\& \quad \operatorname{var}(W(x)) \geq \frac{\left(\frac{\partial}{\partial\theta}\mathbb{E}[W(x)]\right)^2}{-\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right]}$$

Corollary 7.3.10 (Cramer-Rao inequality, iid case)

if $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$

$$\mathrm{Var}_\theta(W(X)) \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(X)]\right)^2}{n\,\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]}$$

Proof:

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]$$

$$= n\, E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]$$

Since $X_1, X_2, \ldots X_n$ iid

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_n(x|\theta)\right)^2\right] = \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log \prod_{i=1}^{n} f(x_i|\theta)\right)^2\right]$$

$$= \mathbb{E}_\theta\left[\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)^2\right]$$

$$= \sum_{i=1}^{n} \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)^2\right]$$

$$+ \sum_{i\neq j} \mathbb{E}_\theta\left[\cancel{\frac{\partial}{\partial\theta} f(x_i|\theta) \frac{\partial}{\partial\theta} f(x_j|\theta)}\right] \to 0 \quad \text{iid}$$

$$= \sum_{i=1}^{n} \mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x_i|\theta)\right)^2\right]$$

$$= n\, E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]$$

<u>Lemma 7.3.11</u> if $f(x|\theta)$ satisfies

$$\frac{d}{d\theta} E_\theta\left[\frac{\partial}{\partial\theta}\log f(x|\theta)\right]$$

$$= \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) f(x|\theta)\right] dx$$

(true for an exponential family)

then

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right]$$

$$= -\mathbb{E}_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right)$$

Example 7.3.12

$$\tau(\lambda)=\lambda \quad , \quad \text{so} \quad \tau(\lambda)'=1$$

$$\mathbb{E}_\lambda\left[\left(\frac{\partial}{\partial\lambda}\log\prod_{i=1}^{n} f(x_i|\lambda)\right)^2\right)$$

$$= -n\mathbb{E}_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log f(x|\lambda)\right)$$

$$= -n\mathbb{E}_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)\right]$$

$$= -n\,\mathbb{E}_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\left(-\lambda + x\log\lambda - \log x!\right)\right\}$$

$$= -n\,\mathbb{E}_\lambda\left[\frac{\partial}{\partial\lambda}\left(-1 + \frac{x}{\lambda}\right)\right]$$

$$= -n\,\mathbb{E}_\lambda\left[-\frac{x}{\lambda^2}\right]$$

$$= \frac{n}{\lambda^2}\cdot\lambda = \frac{n}{\lambda}$$

$$\Rightarrow \boxed{\mathrm{Var}_\lambda W \geq \frac{\lambda}{n}}$$

we know that $\mathrm{Var}_\lambda \bar{x} = \frac{\lambda}{n}$

$\Rightarrow$ $\bar{x}$ is Best unbaised estimator of λ.

Example 7.3.14 (Normal variance Bound)

Let $X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2)$

$$\text{Var}\, W(X) \geq \frac{\left(\frac{d}{d\theta} E[W(X)]\right)^2}{\mathbb{E}\left[\left(\frac{\partial}{\partial\theta} \log f_n(X|\theta)\right)^2\right]}$$

$$\geq \frac{\left(\frac{d}{d\theta} \mathbb{E}[W(X)]\right)^2}{-n\,\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(x|\theta)\right]}$$

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

$$f(x|\theta) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

$$\log f(x|\theta) = -\frac{1}{2}\log 2\pi - \log\sigma - \frac{1}{2\sigma^2}(x-\mu)^2$$

$$\frac{\partial^2}{\partial\theta^2}\log f(x|\theta) = \frac{1}{2\sigma^4} - \frac{(x-\mu)^2}{\sigma^6}$$

$$-n\mathbb{E}\left\{\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right\} = n\mathbb{E}\left[\frac{(x-\mu)^2}{\sigma^6} - \frac{1}{2\sigma^4}\right]$$

$$= n\left[\frac{1}{\sigma^4} - \frac{1}{2\sigma^4}\right]$$

$$\text{Var } W(x) \geq \frac{2\sigma^4}{n}$$

we know that

$$\text{Var } S^2 = \frac{2\sigma^4}{n-1}$$

$\Rightarrow$ S^2 does not attain the Cramer - Rao lower Bound.

* The condition's to attain Cramer - Rao lower Bound are actually quite simple

in Cauchy-Schwarz inequality

$$Cov(X,Y)^2 \leq var(X)\, var(Y)$$

we need to attain equality sign.

$$Cov(X,Y) = var(X)\, var(Y)$$

if $Y = aX + b$; $a \neq 0$

Corollary 7.3.15 (attainment):

Let $X_1, X_2, \ldots, X_n$ iid $f(x|\theta)$, where $f(x|\theta)$ satisfies the condition's of the Cramer-Rao theorem. let $L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta)$ denote the Likelihood function. if $W(X) = W(X_1, \ldots, X_n)$ is any unbaised estimator of $\tau(\theta)$, then $W(X)$ attain's the Cramer-Rao lower Bound $\iff$

$$a(\theta)\left[W(x) - \tau(\theta)\right] = \frac{\partial}{\partial\theta} \log L(\theta|x)$$

for some function $a(\theta)$

Example 7.3.16

$$L(\mu, \sigma^2 | x) = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2} \sum (x-\mu)^2\right\}$$

$$\log L(\mu, \sigma^2 | x) = -\frac{n}{2}\log 2\pi - \frac{n}{2}\log \sigma^2 - \frac{1}{2\sigma^2}\sum (x-\mu)^2$$

$$\frac{\partial}{\partial \sigma^2} \log L(\mu, \sigma^2 | x) = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum (x-\mu)^2$$

$$= \frac{n}{2\sigma^4}\left[\frac{\sum (x_i-\mu)^2}{n} - \sigma^2\right]$$

$$\Rightarrow \quad a(\theta) = \frac{n}{2\sigma^4}$$

$$W(x) = \frac{\sum (x_i-\mu)^2}{n}$$ which depend's on unknown $\mu \Rightarrow$ we cannot attain Crame-Rao lower bound.