Suppose $\underline{X} = X_1, \dots, X_n$ random sample

Then $E[(S(\underline{X}|\theta)^2] = E[(\frac{d}{d\theta} \log f(\underline{X}|\theta)^2] = E[(\frac{d}{d\theta} \log \prod_{i=1}^{n} f(X_i|\theta))^2]$

$$= E[(\sum_{i=1}^{n} \frac{d}{d\theta} \log f(X_i|\theta))^2] = E[(\sum_{i=1}^{n} S(X_i|\theta)^2]$$

$$= E[\sum_{i=1}^{n} S(X_i|\theta)^2] + \underbrace{E[\sum_{i \neq j} S(X_i|\theta) \cdot S(X_j|\theta)]}_{=0} = n E[S(X|\theta)^2]$$

Thereby Corrolar 7.3.10. Cramer-Rao iid.

If in addition to the assumptions in 7.3.9, $X_1, \dots, X_n$ are iid, then $Var[W(\underline{X})] \geq \dfrac{(\frac{d}{d\theta} E[W(\underline{X})])^2}{n E[(\frac{d}{d\theta} \log f(X|\theta))^2]}$

Lemma 7.3.11

Suppose $\frac{d}{d\theta} E[\frac{d}{d\theta} \log f(\underline{X}|\theta)] = \int_{\mathcal{X}} \frac{d}{d\theta} [\frac{d}{d\theta} \log f(\underline{x}|\theta) f(\underline{x}|\theta)] d\underline{x}$

Then $E[(\frac{d}{d\theta} \log f(\underline{X}|\theta))^2] = - E[\frac{\partial^2}{\partial\theta^2} \log f(\underline{X}|\theta)]$

or $E[S(\underline{X}|\theta)^2] = - E[\frac{d}{d\theta} S(\underline{X}|\theta)]$

Proof. $0 = \int_{\underline{\mathcal{X}}} (\frac{d}{d\theta} \log f(\underline{x}|\theta)) f(\underline{x}|\theta) d\underline{x} \Leftrightarrow 0 = \int_{\underline{\mathcal{X}}} \frac{d}{d\theta} [\frac{d}{d\theta} \log f(\underline{x}|\theta) f(\underline{x}|\theta)] d\underline{x}$

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

$$\Longleftrightarrow \int_{\underline{\mathcal{X}}} \frac{d^2}{d\theta^2} \log f(\underline{x}|\theta) f(\underline{x}|\theta) d\underline{x} + \int_{\underline{\mathcal{X}}} (\frac{d}{d\theta} \log f(x|\theta))^2 f(x|\theta) d\underline{x} = 0$$

$$\Longleftrightarrow E[\frac{d}{d\theta} S(\underline{X}|\theta)] + E[(S(\underline{X}|\theta))^2] = 0$$

Some Examples 7.3.12

Poisson distribution $X_1, \ldots, X_n$ iid

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

$\tau(\lambda) = \lambda \Rightarrow \tau'(\lambda) = 1$

$f(x|\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \Rightarrow \log f(x|\lambda) = x \log \lambda - \lambda - \log(x!)$

$S(x|\lambda) = \frac{d}{d\lambda} \log f(x|\lambda) = \frac{x}{\lambda} - 1$, $\frac{d}{d\lambda} S(x|\lambda) = -\frac{x}{\lambda^2}$

$-n E\left[-\frac{X}{\lambda^2}\right] = \frac{n}{\lambda} \Rightarrow Var(\bar{X}) \geq \frac{1}{\frac{n}{\lambda}} = \frac{\lambda}{n} = Var(\bar{X})$. No need to check S^2.

7.3.13 $X_1, \ldots, X_n$ iid.

$$f(x|\theta) = \begin{cases} \frac{1}{\theta}, & 0 < x < \theta \\ 0 & \text{otherwise} \end{cases} = \begin{cases} e^{-\log \theta}, & 0 < x < \theta, \quad \theta > 0 \\ 0 & \text{otherwise.} \end{cases}$$

$S(x|\theta) = -\frac{1}{\theta}$. $E[S(x|\theta)^2] = \frac{1}{\theta^2}$ indicates that for an unbiased estimator W of θ we have $Var[W] \geq \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n}$

$Y = \max_i X_i$ is sufficient. $F_Y(y) = \left(\frac{y}{\theta}\right)^n$ and $f_Y(y) = \frac{n y^{n-1}}{\theta^n}$

and $E[Y] = \int_0^\theta \frac{n y^n}{\theta^n} dy = \frac{n}{\theta^n}\left[\frac{y^{n+1}}{n+1}\right]_0^\theta = \frac{n}{n+1}\theta$

$\Rightarrow \frac{n+1}{n} Y$ is an unbiased estimator of θ.

$$Var\left[\frac{n+1}{n} Y\right] = \frac{(n+1)^2}{n^2} Var[Y] = \frac{(n+1)^2}{n^2}\left[\int_0^\theta \frac{n y^{n+1}}{\theta^n} dy - \left(\frac{n}{n+1}\theta\right)^2\right]$$

$$= \frac{(n+1)^2}{n^2}\left[\frac{n}{n+2}\theta^2 - \left(\frac{n\theta}{n+1}\right)^2\right] = \frac{\theta^2}{n(n+2)} < \frac{\theta^2}{n}$$

Why $\frac{d}{d\theta}\int_0^\theta h(x) f(x|\theta)\, dx = \frac{d}{d\theta}\int_0^\theta h(x)\frac{1}{\theta}\, dx = \frac{h(\theta)}{\theta} - \int_0^\theta h(x)\frac{1}{\theta^2}\, dx$

2.4.1. Leibnitz rule

$\neq \quad - \int_0^\theta h(x)\frac{1}{\theta^2}\, dx$ unless $h(x) = 0$. (a.e).

7.3.2. Sufficiency and Unbiasedness

Theorem 7.3.17

Rao-Blackwell. Let W be an ~~a~~ unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistics for θ. Define $\phi(T) = E[W|T]$. Then $E[\phi(T)] = \tau(\theta)$ and $Var[\phi(T)] \leq Var[W]$, $\forall\theta$, that is $\phi(T)$ is a uniformly better unbiased estimator for $\tau(\theta)$.

Proof. $\tau(\theta) = E[W] = E[E[W|T]] = E[\phi(T)]$

$$Var[W] = Var[E[W|T]] + E[Var[W|T]]$$

$$= Var[\phi(T)] + E[Var[W|T]] \geq Var[\phi(T)]$$

Since for a sufficient statistic the distribution of $W|T$ does not depend on θ, $\phi(T)$ is only a function of the sample $X_1, \ldots, X_n$.

The multiparameter case $\underline{\theta} = [\theta_1, \dots, \theta_K]$

Define $$\underline{S}(\underline{X}|\underline{\theta}) = \begin{bmatrix} \frac{d}{d\theta_1} \log f(\underline{x}|\underline{\theta}) \\ \vdots \\ \frac{d}{d\theta_k} \log f(\underline{x}|\underline{\theta}) \end{bmatrix} = \nabla \log f(\underline{x}|\underline{\theta})$$

Define $\underline{I}(\underline{\theta}) \overset{\text{def}}{=} \text{Cov}[\underline{S}(\underline{x}|\underline{\theta})]$. We have $E[\underline{S}(\underline{x}|\underline{\theta})] = \underline{0}$

and $\underline{I}(\underline{\theta}) = E[\underline{S}(\underline{X}|\underline{\theta}) \cdot \underline{S}(\underline{x}|\underline{\theta})^T] = -E\,\underline{H}(\underline{X}, \underline{\theta})$

where $h_{ij} = \dfrac{\partial^2}{\partial\theta_i \partial\theta_j} \log f(\underline{x}|\underline{\theta})$

Let $\tau = \tau(\underline{\theta})$ be univariate and let $\nabla \tau(\underline{\theta}) = \begin{bmatrix} \frac{d}{d\theta_1} \tau(\underline{\theta}) \\ \vdots \\ \frac{d}{d\theta_k} \tau(\underline{\theta}) \end{bmatrix}$

For an estimator $W(\underline{X})$ with $E[W(\underline{X})] = \tau(\underline{\theta})$

we have under similar regularity conditions as in the univariate case that

$$\text{Var}[W(\underline{X})] \geq (\nabla \tau(\underline{\theta}))' (\underline{I}(\underline{\theta}))^{-1} (\nabla \tau(\underline{\theta}))$$

Sketch of proof

$$\frac{\partial}{d\theta_i} \tau(\underline{\theta}) = \frac{\partial}{\partial\theta_i} \int_{\underline{x}} W(\underline{x}) f(\underline{x}, \underline{\theta})\, d\underline{x} = \int_{\underline{x}} W(\underline{x}) \left[\frac{\partial}{d\theta_i} f(\underline{x}, \underline{\theta})\right] d\underline{x}$$

$$= \int_{\underline{x}} W(\underline{x}) \left(\frac{\partial}{d\theta_i} \log f(\underline{x}|\underline{\theta})\right) f(\underline{x}|\underline{\theta})\, d\underline{x} = E[W(\underline{x})\, S_i(\underline{x}|\underline{\theta})]$$

$$\Rightarrow \nabla \tau(\underline{\theta}) = E[W(\underline{X})\, \underline{S}(\underline{x}|\underline{\theta})]$$

Define $U(\underline{X}, \underline{\theta}) = [\nabla \tau(\underline{\theta})]' (\underline{I}(\underline{\theta}))^{-1} \underline{S}(\underline{X}, \underline{\theta})]$ is univariate and $Var[u(\underset{\sim}{X}, \underset{\sim}{\theta})] = [\nabla \tau(\underset{\sim}{\theta})]'(\underset{\sim}{I}(\underset{\sim}{\theta}))^{-1} \cdot \underset{\sim}{I}(\underset{\sim}{\theta}) \cdot (\underset{\sim}{I}(\underset{\sim}{\theta}))^{-1} \nabla \tau(\underset{\sim}{\theta})]$

$= [\nabla \tau(\underset{\sim}{\theta})]'(\underset{\sim}{I}(\underset{\sim}{\theta}))^{-1}[\nabla \tau(\underset{\sim}{\theta})]$

$Cov[W(\underset{\sim}{X}), u(\underset{\sim}{X}, \underset{\sim}{\theta})] = [\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})]'[\underset{\sim}{I}(\underset{\sim}{\theta})]^{-1} E[\underset{\sim}{S}(\underset{\sim}{X}, \underset{\sim}{\theta}) W(\underset{\sim}{X})]$

$= [\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})]'[\underset{\sim}{I}(\underset{\sim}{\theta})]^{-1}[\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})]$

Cauchy Schwarstz implies.

$[[\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})]'(\underset{\sim}{I}(\underset{\sim}{\theta}))^{-1}(\nabla \tau(\underset{\sim}{\theta}))]^2 \le [\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})]'[\underset{\sim}{I}(\underset{\sim}{\theta})]^{-1}[\nabla \underset{\sim}{\tau}(\underset{\sim}{\theta})] \cdot Var[W(\underset{\sim}{X})]$