Definition: An unbiased estimator $T: \mathcal{X} \longrightarrow E \in \mathbb{R}$ of $g(\vartheta)$ is called **efficient** if $\forall \vartheta \in \Theta$:

$$Var_\vartheta(T) = \nabla_\vartheta g(\vartheta) \left(I(\vartheta)\right)^{-1} \nabla_\vartheta g(\vartheta)^T$$

Theorem: (Cramer-Rao) Let T be an unbiased estimator of $g(\vartheta) \in \mathbb{R}$ with $\mathbb{E}_\vartheta(T^2 < \infty)$, $\forall \vartheta \in \Theta$. Then assuming I is invertible $\forall \vartheta$:

$$Var_\vartheta(T) \geq \nabla_\vartheta g(\vartheta)^T \left[I(\vartheta)\right]^{-1} \cdot \nabla_\vartheta g(\vartheta)$$

Note: if $\Theta = \mathbb{R}$ $\quad Var_\vartheta(T) \geq \left(\partial_\vartheta g(\vartheta)\right)^2 / I(\vartheta)$

if $g(\vartheta) = \vartheta$ $\quad \partial_\vartheta g(\vartheta) = 1$

$\Rightarrow Var_\vartheta(T) \geq \left(I(\vartheta)\right)^{-1}$ $\qquad \lrcorner$

Proof: We assume $g(\vartheta) = \mathbb{E}_\vartheta[T(x)]$

$$\nabla_\vartheta g(\vartheta) = \nabla_\vartheta \mathbb{E}_\vartheta[T(x)] = \nabla_\vartheta \int T(x)\, p_\vartheta(x)\, \mu(dx)$$

$$= \int T(x) \nabla_\theta p_\theta(x) \mu(dx)$$

$$= \int T(x) \nabla_\theta p_\theta(x) \cdot \frac{1}{p_\theta(x)} \cdot p_\theta(x) \mu(dx)$$

$$= \int T(x) \nabla_\theta \log p_\theta(x) \cdot p_\theta(x) \mu(dx)$$

$$= \mathbb{E}_\theta [T(x) \nabla_\theta \log p_\theta(x)]$$

$$\mathrm{Cov}_\theta(T(x), \nabla_\theta \log p_\theta(x)) = \qquad (\mathbb{E}[X \cdot Y] - \mathbb{E}[X] \cdot \mathbb{E}[Y])$$

$$= \mathbb{E}_\theta [T(x) \cdot \nabla_\theta \log p_\theta(x)] - \mathbb{E}_\theta[T(x)] \cdot \underbrace{\mathbb{E}_\theta [\nabla_\theta \log p_\theta(x)]}_{=0}$$

$$= \mathbb{E}_\theta \left(\nabla_\theta \log p_\theta(x) \underbrace{(T(x) - \mathbb{E}_\theta(T(x)))}_{\text{centered estimator}} \right)$$

$$\langle x, \nabla_\theta \mathbb{E}_\theta(T(x)) \rangle^2 =$$

$$= \mathbb{E}_\theta \left(\langle x, \nabla_\theta \log p_\theta(x) \rangle \cdot (T(x) - \mathbb{E}_\theta(T(x))) \right)^2$$

$$\nabla_\theta \mathbb{E}_\theta[T(x)] = \mathbb{E}[T(x) - \mathbb{E}(T(x))) \cdot \nabla_\theta \log p_\theta]$$

By Cauchy-Schwartz inequality

$$(\mathbb{E}[fg])^2 \leq \mathbb{E}[f^2] \cdot \mathbb{E}[g^2]$$

$$f(x) = \langle x, \nabla_\theta \log p_\theta(x) \rangle$$

$$g(x) = T(x) - \mathbb{E}_\theta[T(x)]$$

$$\left(\langle x, \nabla_\theta \mathbb{E}_\theta[T(x)] \rangle\right)^2 \leq \mathrm{var}_\theta(T) \cdot$$

$$\cdot \mathbb{E}[(\langle x, \nabla_\theta \log p_\theta(x) \rangle)^2]$$

$$\Rightarrow \mathrm{Var}_\theta(T) \geq \frac{\langle x, \nabla_\theta \mathbb{E}_\theta[T(x)] \rangle^2}{\underbrace{\mathbb{E}_\theta[(\langle x, \nabla_\theta \log p_\theta(x) \rangle)^2]}_{*}}$$

But note that $* = x^T \underbrace{I(\vartheta)}_{\text{fisher Information matrix}} \cdot x$

In order to get the tightest lower bound

we maximize the RHS over all directions $x \in \mathbb{R}^d$

$$Var_\theta(T) \geq \sup_x \frac{|\langle x, \nabla_\theta g(\theta)\rangle|^2}{x^T \cdot I(\theta) \cdot x}$$

「 Lemma let $a \in \mathbb{R}^k$, $A \in \mathbb{R}^{k \times k}$ symmetric positive definite then :

$$\sup_{x \in \mathbb{R}^k} \frac{\langle x, a\rangle^2}{\langle Ax, x\rangle} = \langle A^{-1}a, a\rangle$$

Proof: let $BB = A$, B symmetric:

$$\frac{\langle x, a\rangle}{\langle Ax, x\rangle} = \frac{\langle B^{-1}y, a\rangle}{\langle AB^{-1}y, B^{-1}y\rangle} = \frac{\langle y, B^{-1}a\rangle}{\|y\|^2}$$

$$= \left(\frac{\langle y, b\rangle}{\|y\|}\right)^2 = \|b\|^2 = \langle (B^2)^{-1}a, a\rangle$$

↑ max in $y = b$ 」

Applying this lemma in the proof

$$\sup_x \frac{(\langle x, \nabla_\theta g(\theta) \rangle)^2}{x^T I(\theta) x} = \nabla_\theta g(\theta)^T (I(\theta))^{-1} \cdot \nabla_\theta g(\theta)$$

Remark: Project the estimator's sensitivity into the direction which carries the most information (according to Fisher Inform matrix)

Then optimize over all x

Definition: we call $\nabla_\theta g(\theta)^T (I(\theta))^{-1} \cdot \nabla_\theta g(\theta)$ the Cramer-Rao Lower Bound for estimating $g(\theta)$ (CRLB)

Question: UMVU estimator $\overset{?}{\neq}$ CRLB

Example: (Poisson)

$$X_i \overset{\text{i.i.d}}{\sim} \text{Pois}(\theta) \qquad i = 1, \dots, n$$

$$\sum x_i \sim \text{Pois}(n\theta) \qquad \bar{X} : \text{sufficient statistic}$$

for sample

$$\log p_\theta(x) = -n\theta + \sum x_i \log\theta - \sum \log x_i!$$

for one observation

$$\log p_\theta(x) = -\theta + x\log\theta - \log x!$$

Score function

$$S_\theta(x) = \nabla_\theta \log p_\theta(x) = -1 + \frac{x}{\theta}$$

Fisher Information for Sample:

$$I_1(\theta) = \text{Var}_\theta(S_\theta(x)) = \text{Var}_\theta\left(-1 + \frac{x}{\theta}\right) = \frac{1}{\theta^2}\text{Var}_\theta(x)$$

$$= \frac{1}{\theta}$$

$$I_n(\theta) = n I_1(\theta) = \frac{n}{\theta}$$

Estimate $g(\vartheta) = \vartheta$

Let $\hat{\vartheta} = \frac{1}{n} \sum x_j = \bar{x}$

$$Var(\hat{\vartheta}) = \frac{1}{n} \cdot \vartheta$$

$g(\vartheta) = \vartheta \implies g'(\vartheta) = 1$

$$CRLB = \frac{(g'(\vartheta))^2}{I_n(\vartheta)} = \frac{\vartheta}{n}$$

Estimator reaches CRLB it is efficient + unbiased so this estimator $g(\vartheta) = \vartheta$ is UMVU

CRLB $\implies$ UMVU.

Now consider

$$T = \left(\frac{n-1}{n}\right)^{\sum x_j}$$

$$E_\vartheta[T] = e^{-n\vartheta\left(1 - \frac{n-1}{n}\right)} = e^{-\vartheta} \quad (\text{MGF})$$

This is an unbiased estimator for

$g(\vartheta) = e^{-\vartheta}$

Compute the variance of T

$$Var_\vartheta(T) = E_\vartheta(T^2) - (E_\vartheta(T))^2$$

$$T^2 = \left(\frac{n-1}{n}\right)^{2S} = \left(\frac{(n-1)^2}{n^2}\right)^S$$

$$E[T^2] = \sum_{j=0}^{\infty} \left(\frac{(n-1)^2}{n^2}\right)^j \cdot \frac{(n\vartheta)^j}{j!} \cdot e^{-n\vartheta}$$

$$= \exp\left(n\vartheta\left(\frac{(n-1)^2}{n^2} - 1\right)\right) = e^{-n\vartheta} \cdot e^{\vartheta \cdot \frac{2n-1}{n}}$$

$$Var_\vartheta(T) = e^{-2\vartheta}\left(\exp\left(\frac{(n-1)^2}{n} \cdot \vartheta\right) - 1\right)$$

$$= e^{-2\vartheta} \cdot \frac{\vartheta}{n} = \frac{g(\vartheta)^2 \cdot \vartheta}{n}$$

CRLB $\quad g'(\theta) = -e^{-\theta} = -g(\theta)$

$$\Rightarrow \quad (g'(\theta))^2 = g(\theta)^2$$

$$\Rightarrow \quad CRLB = \frac{(g'(\theta))^2}{I_n(\theta)} = \frac{g(\theta)^2}{n/\theta} = \frac{g(\theta)^2\theta}{n}$$

this is not an UMVU estimator

even though it reaches CRLB

and it is efficient