Defn: <u>Uniformly Minimum - Variance Unbiased Estimator</u>

(UMVUE)

We call $\hat{\theta}^*$ the UMVUE of $\tau(\theta)$ if

e.g. θ^2 or $\sqrt{\theta}$ or $\log\theta$

① (unbiased)

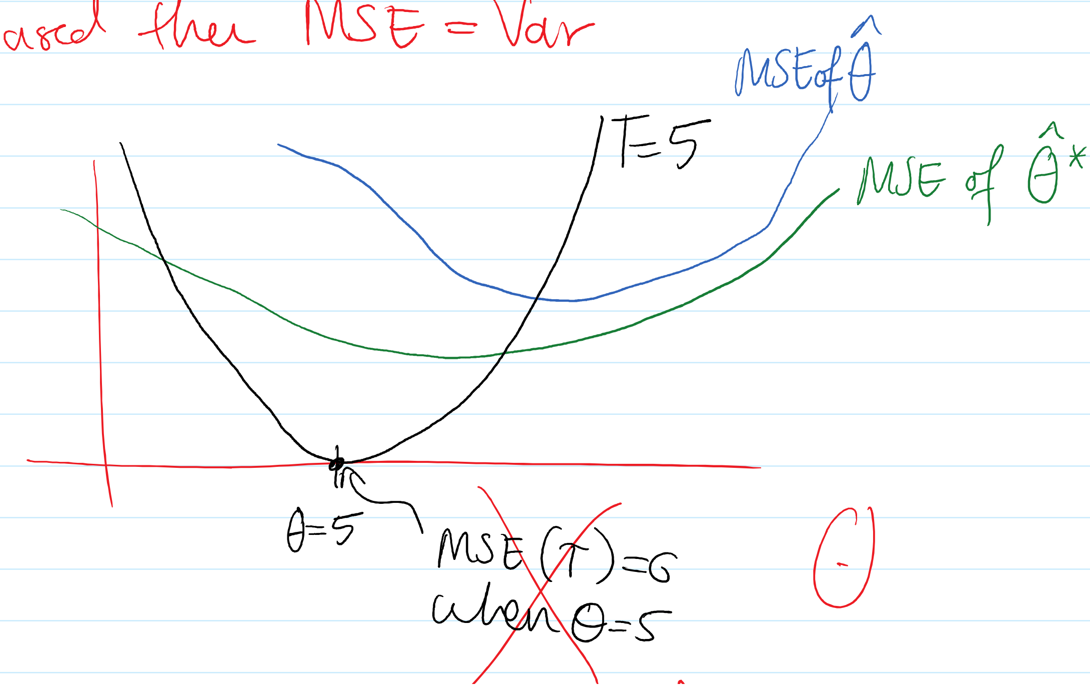
$$E[\hat{\theta}^*] = \tau(\theta)$$

② (minimum variance)

$$Var_\theta(\hat{\theta}^*) \le Var_\theta(\hat{\theta}) \quad \forall\, \theta \in \Theta$$

for any other unbiased estimator $\hat{\theta}$ of $\tau(\theta)$

If unbiased then MSE = Var



$\cancel{\text{when}}\ \theta = 5$

Not allowed

<u>Defn: Score</u> Recall $\underset{\sim}{X}$ is random

and $\underset{\sim}{x}$ is not (just in $\mathbb{R}^N$)

If $X_n \overset{iid}{\sim} f_\theta$ where $\theta \in \Theta$ then the score

$$\boxed{S_\theta(\underset{\sim}{X}) = \text{"}\frac{\partial \ell}{\partial \theta}\text{"} = \frac{\partial \log f_\theta}{\partial \theta}(\underset{\sim}{X})}$$

$\left[\frac{\partial f_\theta}{\partial \theta}\right](\underset{\sim}{X})$

notice difference btw $f_\theta(\underset{\sim}{x})$ and $f_\theta(\underset{\sim}{X})$

<u>Ex.</u> $X_n \overset{iid}{\sim} Exp(\lambda)$ so that $f_\lambda(x) = \lambda e^{-\lambda x} \underbrace{\mathbb{1}(x>0)}$

doesn't depend on λ

so

$$L(\lambda) = f_\lambda(\underset{\sim}{x}) = \prod_{n=1}^{N} \lambda e^{-\lambda x_n} \mathbb{1}(x_n > 0)$$

$$= \lambda^N e^{-\lambda \sum_{n=1}^{N} x_n} \mathbb{1}(x_{(1)} > 0)$$

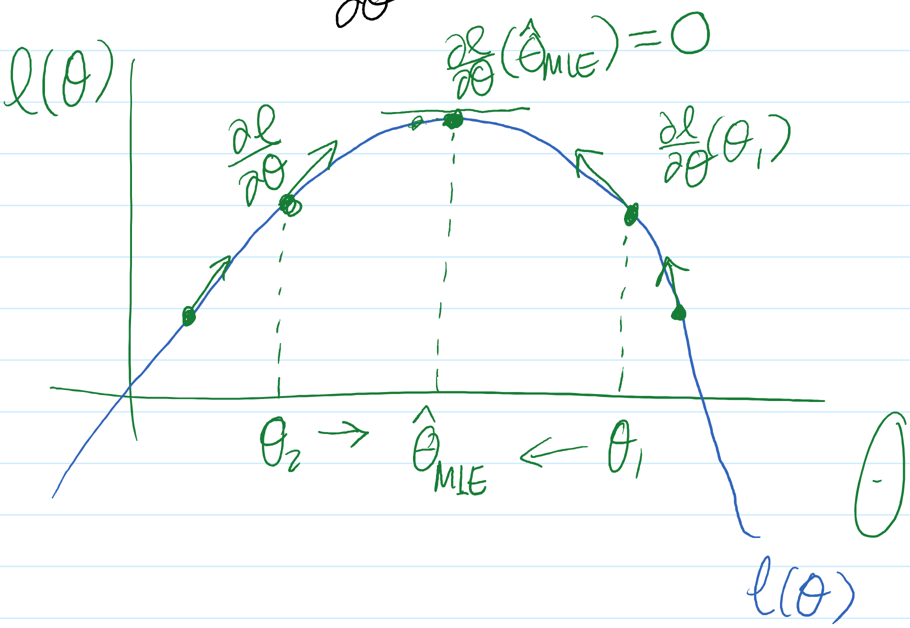
optional

$$\ell(\lambda) = N \log \lambda - \lambda \sum_{n=1}^{N} x_n + \log \mathbb{1}(x_{(1)} > 0)$$

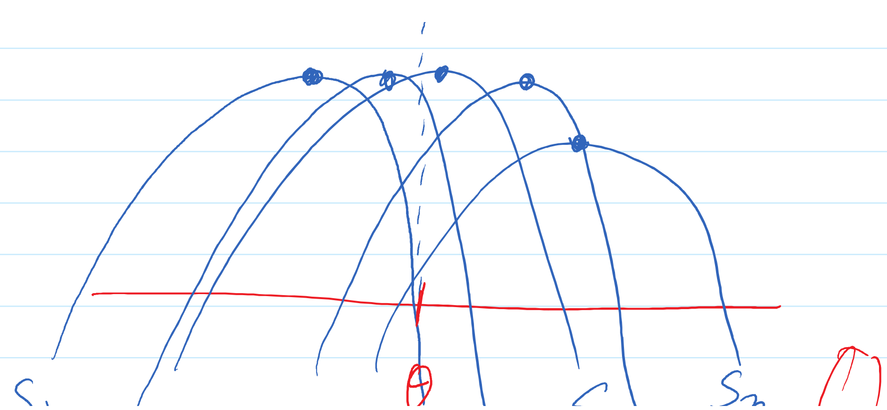
$$\boxed{\frac{\partial \ell}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=1}^{N} x_n} \leftarrow \text{deterministic}$$

$$\boxed{S_{\lambda}(X) = \frac{N}{\lambda} - \sum_{n=1}^{N} X_n} = \text{"}\frac{\partial \ell}{\partial \lambda}\text{"} \longleftarrow \text{random}$$

For the <u>deterministic version</u>
what does $\frac{\partial \ell}{\partial \theta}$ tell us?



The <u>score</u>: Fact $E[S] = 0$.



<u>Proof</u>: $\boxed{E[S_\theta] = 0.}$

$$E_\theta[S_\theta] = \int S_\theta(\underset{\sim}{x}) f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x}$$

lazy notation $\int\int \cdots \int \cdots\cdots dx_1 dx_2 \cdots dx_N$

$$= \int \frac{\partial \ell}{\partial \theta} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x}$$

$$= \int \frac{\frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})}{\cancel{f_\theta(\underset{\sim}{x})}} \cancel{f_\theta(\underset{\sim}{x})}\, d\underset{\sim}{x}$$

Aside:

$$\frac{\partial \ell}{\partial \theta} = \frac{\partial}{\partial\theta} \log f_\theta(\underset{\sim}{x}) = \frac{\frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})}{f_\theta(\underset{\sim}{x})}$$

$$\frac{d}{dx} \log u(x) = \frac{u'(x)}{u(x)}$$

$$= \int \frac{\partial}{\partial\theta} f_\theta(\underset{\sim}{x})\, d\underset{\sim}{x} \quad (*)$$

$(*)$ We have enough "regularity" / "niceness" to interchange integration/differentiation.

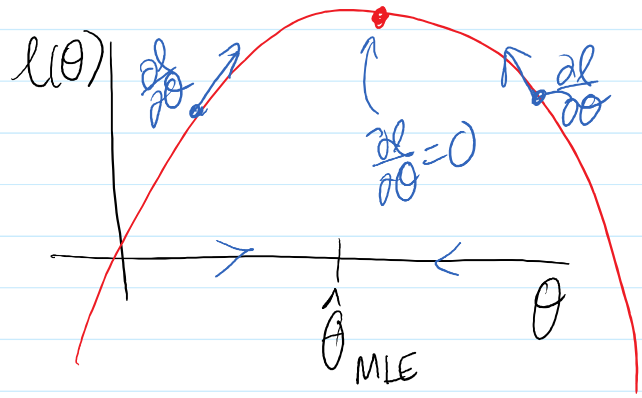
$$= \frac{\partial}{\partial\theta} \underbrace{\int f_\theta(\underset{\sim}{x})\, dx}_{\int \text{PDF} = 1}$$

$$= \frac{\partial}{\partial\theta}(1) = 0$$

when is this a valid move?

(1) Exponential family (f_θ)

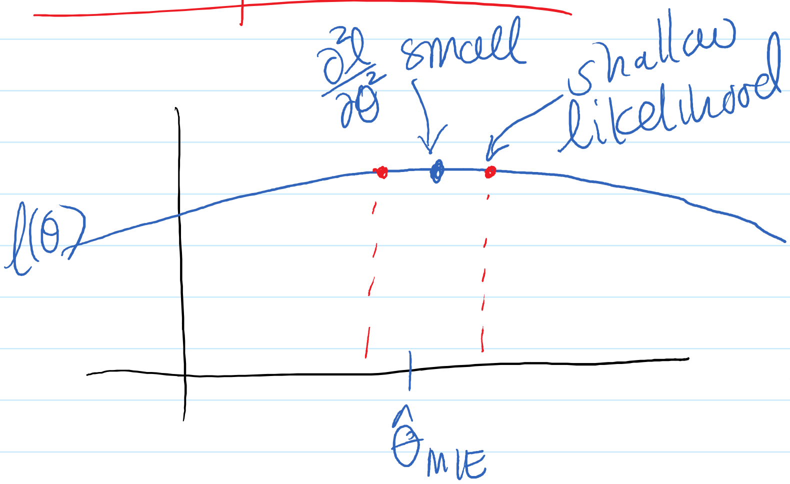
② Not when the support of f_θ depends on θ (e.g. $\mathrm{Unif}(0,\theta)$.)

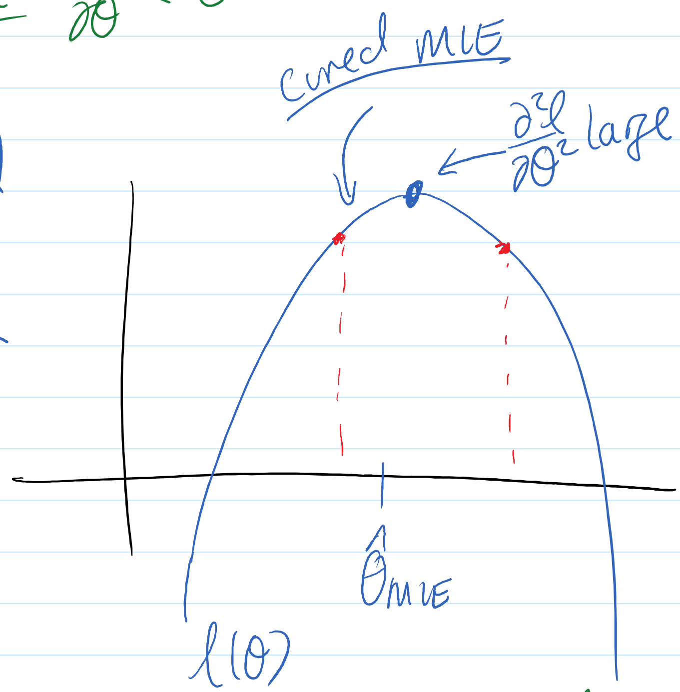


What about $\frac{\partial^2 \ell}{\partial \theta^2}$?

Note: $\frac{\partial^2 \ell}{\partial \theta} < 0$

Two possibilities:





Weakly prefer $\hat{\theta}_{MLE}$

Strongly prefer $\hat{\theta}_{MLE}$

Theorem:

$$\boxed{\mathrm{Var}(\underbrace{S_\theta}_{\text{“}\frac{\partial \ell}{\partial \theta}(\underline{X})\text{”}}) \overset{1}{=} E[S_\theta^2] \overset{2}{=} -E\left[\frac{\partial^2 \ell}{\partial \theta^2}(\underline{X})\right]}$$

$$\boxed{Var(S_\theta) \overset{①}{=} E[S_\theta^2] \overset{②}{=} -E\left[\frac{\partial^2 \ell}{\partial\theta^2}(\underline{x})\right]}$$

Pf. ① Note $E[S_\theta] = 0$ so $Var(S_\theta) = E[S_\theta^2] - \cancel{E[S_\theta]^2}$ 0

② Another way: " $S_\theta = \frac{\partial \ell}{\partial\theta}$ "

$$\underbrace{E\left[\left(\frac{\partial\ell}{\partial\theta}\right)^2\right]}_{>0} = \underbrace{\underbrace{-}_{<0}\ \underbrace{E\left[\frac{\partial^2\ell}{\partial\theta^2}\right]}_{<0}}_{>0}$$

Pf.

$$\begin{aligned}
0 &= \frac{\partial}{\partial\theta} 0 = \frac{\partial}{\partial\theta} E[S_\theta] \\
&= \frac{\partial}{\partial\theta} \int \frac{\partial\ell}{\partial\theta} f_\theta(\underline{x})\, d\underline{x} \\
&= \int \frac{\partial}{\partial\theta}\left[\frac{\partial\ell}{\partial\theta} f_\theta(\underline{x})\right] d\underline{x} \quad (*) \text{ regularity} \\
&= \int \left[\frac{\partial}{\partial\theta}\frac{\partial\ell}{\partial\theta} f_\theta(\underline{x}) + \frac{\partial\ell}{\partial\theta}\frac{\partial}{\partial\theta} f_\theta(\underline{x})\right] dx \\
&= \int \left[\frac{\partial^2\ell}{\partial\theta^2} f_\theta(\underline{x}) + \frac{\partial\ell}{\partial\theta}\frac{\partial\ell}{\partial\theta} f_\theta(\underline{x})\right] d\underline{x}
\end{aligned}$$

$$\frac{\partial\ell}{\partial\theta} = \frac{\frac{\partial}{\partial\theta} f_\theta}{f_\theta} \iff \frac{\partial}{\partial\theta} f_\theta = \frac{\partial\ell}{\partial\theta} f_\theta$$

$$0 = \underbrace{\int \frac{\partial^2\ell}{\partial\theta^2} f_\theta(\underline{x})\, d\underline{x}}_{E\left[\frac{\partial^2\ell}{\partial\theta^2}\right]} + \underbrace{\int \left(\frac{\partial\ell}{\partial\theta}\right)^2 f_\theta(\underline{x})\, d\underline{x}}_{E\left[\left(\frac{\partial\ell}{\partial\theta}\right)^2\right]}$$

$$\mathbb{E}\left[\frac{\partial \ell}{\partial \theta}\right] + \mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right]$$

So $$\boxed{\mathbb{E}\left[\left(\frac{\partial \ell}{\partial \theta}\right)^2\right] = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \theta^2}\right]}$$

$$\mathbb{E}[S_\theta^2] = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right]$$

Defn: Fisher Information If $X \sim f_\theta$ (N=1)
we define the Fisher Information for θ contained in X as

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right]$$

If I have N samples $X_n \overset{iid}{\sim} f_\theta$ then the Fisher Information contained about θ in these N samples is

$$I_N(\theta) = N\, I(\theta).$$

pf: $$I_N(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log f_\theta(X)\right]$$
$$= -\mathbb{E}\left[\frac{\partial^2}{\partial \theta^2} \log \prod_{n=1}^{N} f_\theta(X_n)\right]$$

$$= -E\left[\frac{\partial^2}{\partial\theta^2}\sum_{n=1}^{N}\log f_\theta(X_n)\right]$$

$$= \sum_{n=1}^{N}\underbrace{-E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X_n)\right]}_{\text{Info in each } I(\theta)}$$

$$= NI(\theta).$$

Ex. $X_n \overset{iid}{\sim} \mathrm{Pois}(\lambda) \longrightarrow f_\lambda(x) = \frac{\lambda^x e^{-\lambda}}{x!}$

① find $\log f_\theta$

$$\log f_\lambda(x) = x\log\lambda - \lambda - \log(x!)$$

② find $\frac{\partial^2}{\partial\theta^2}\log f_\theta(x)$

$$\to \frac{\partial}{\partial\lambda}\log f_\lambda(x) = \frac{x}{\lambda} - 1$$

$$\to \frac{\partial^2}{\partial\lambda^2}\log f_\lambda(x) = \frac{-x}{\lambda^2}$$

③ Find $-E\left[\frac{\partial^2 \ell}{\partial\theta^2}\right] = -E\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right] = I(\theta)$

$$-E\left[-\frac{X}{\lambda^2}\right] = \frac{1}{\lambda^2}E[X] = \frac{1}{\lambda^2}\lambda = \frac{1}{\lambda}$$

$$-\mathbb{E}\left[-\frac{X}{\lambda^2}\right] = \frac{1}{\lambda^2}\mathbb{E}[X] = \frac{1}{\lambda^2}\lambda = \boxed{1/\lambda} .$$

④ $\underline{I_N(\theta)}$

$$\boxed{I_N(\lambda) = N/\lambda} .$$