3. $X_1, \cdots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta)$ $\theta > 0$. Find $\hat{\theta}_{MLE}$.

$$L(\theta) = f(X_1, \cdots, X_n; \theta) \overset{iid}{=} f(X_1; \theta) \cdots f(X_n; \theta)$$

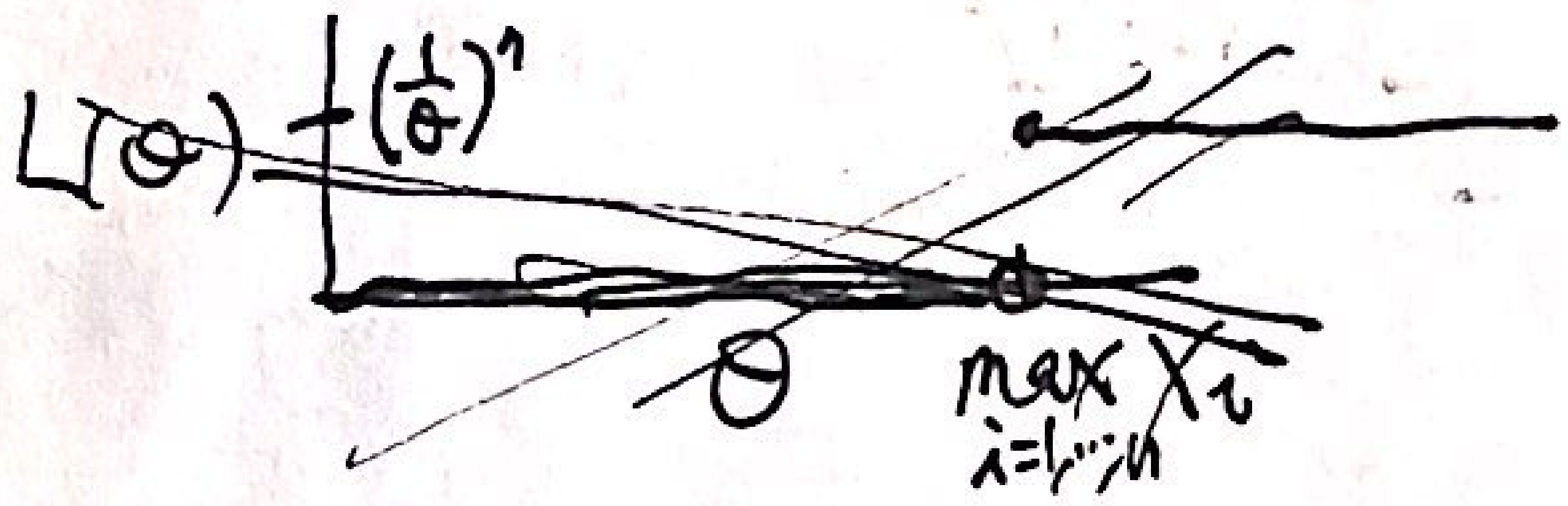
$$f(X_i; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } X_i \in [0, \theta] \\ 0 & \text{otherwise} \end{cases} = \frac{1}{\theta} \cdot I(X_i \le \theta)$$

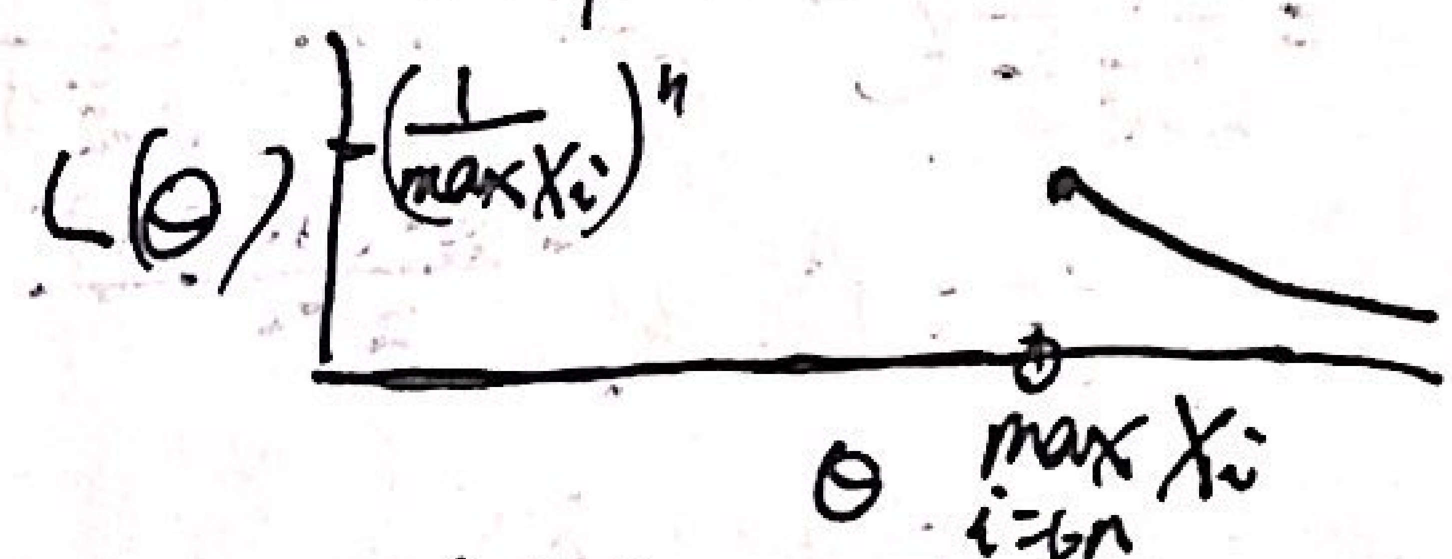
$$\begin{aligned} L(\theta) &= \frac{1}{\theta} \cdot I(X_1 \le \theta) \cdots \frac{1}{\theta} I(X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I(X_1 \le \theta) \cdots I(X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I(X_1 \le \theta, \cdots, X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I(\max_{i=1,\cdots,n} X_i \le \theta) \end{aligned}$$

$\Rightarrow$ not differentiable with respect to θ

$\hat{\theta}_{MOM} = 2\overline{X_n}$

$\hat{\theta}_{MLE} = \max_{i=1,\cdots,n} X_i$





$$\Rightarrow \arg\max_{\theta} L(\theta) = \max_{i=1,\cdots,n} X_i = \max\{X_1, \cdots, X_n\}$$

Properties of $\hat{\theta}$ (estimator of θ)

① Bias: $\text{bias}(\hat{\theta}) = E[\hat{\theta}] - \theta$

② variance: $\text{var}[\hat{\theta}] = E[(\hat{\theta} - E[\hat{\theta}])^2]$ (by def)

③ mean-squared error (MSE): $\text{MSE}(\hat{\theta}) = E[(\hat{\theta} - \theta)^2]$ "how well does $\hat{\theta}$ concentrate on θ"

$$\text{MSE}(\hat{\theta}) = (\text{bias}(\hat{\theta}))^2 + \text{var}(\hat{\theta})$$

Two estimators $\hat{\theta}_1, \hat{\theta}_2$: if $\text{MSE}(\hat{\theta}_1) = \text{MSE}(\hat{\theta}_2)$, then

$$\text{bias}(\hat{\theta}_1) < \text{bias}(\hat{\theta}_2) \Leftrightarrow \text{var}(\hat{\theta}_1) > \text{var}(\hat{\theta}_2)$$

"bias - variance tradeoff"

Cramer-Rao Inequality (Theorem): the minimum variance of unbiased $\hat{\theta}$

Let $X_1, \cdots, X_n \overset{iid}{\sim}$ pdf $f(\cdot; \theta)$. Let $\hat{\theta} = g(X_1, \cdots, X_n)$ be an unbiased estimator of θ. Then under smoothness conditions of $f(\cdot; \theta)$,

$$\text{var}(\hat{\theta}) \ge \frac{1}{n I(\theta)} \quad \text{where } I(\theta) = \text{Fisher information}$$

Recall: $S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \Rightarrow E[S_n^2] = Var(X_i) = \sigma^2$ unbiased

$$\frac{n-1}{n} S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \Rightarrow E\left[\frac{n-1}{n}S_n^2\right] = \frac{n-1}{n}E[S_n^2] = \frac{n-1}{n}\sigma^2$$

$$\Rightarrow bias\left(\frac{n-1}{n}S_n^2\right) = E\left[\frac{n-1}{n}S_n^2\right] - \sigma^2$$

$$= \frac{n-1}{n}\sigma^2 - \sigma^2 = -\frac{1}{n}\sigma^2 \xrightarrow{n\to\infty} 0$$

So $\frac{n-1}{n}S_n^2$ is biased but <u>asymptotically</u> unbiased.

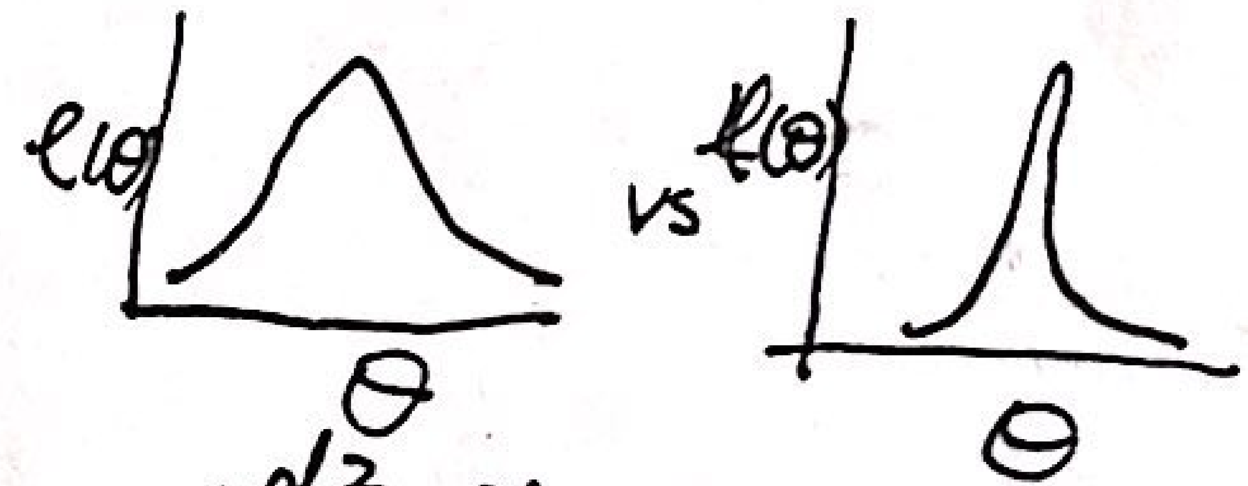
Fisher information: based on the log-likelihood of θ w/ 1 obs

$\ell(\theta) = \log f(X;\theta)$ $\quad$ f is the pdf or pmf of r.v. X

Fisher information: $I(\theta) = \underset{\text{w.r.t. } X}{E}\left[\left(\frac{d}{d\theta}\ell(\theta)\right)^2\right]$ $\quad$ θ is a constant

$$= E_X\left[\left(\frac{d}{d\theta}\log f(X;\theta)\right)^2\right]$$

Alternative def: $I(\theta) = -E\left[\frac{d^2}{d\theta^2}\ell(\theta)\right]$



$-\frac{d^2}{d\theta^2}\ell(\theta)$ is bigger on the right

Pf: Since $\int f(x;\theta)\,dx = 1$

$$\frac{d}{d\theta}\int f(x;\theta)\,dx = \frac{d}{d\theta}1 = 0$$

Because of the smoothness conditions of $f(\cdot;\theta)$,

$$\frac{d}{d\theta}\int f(x;\theta)\,dx = \int \frac{d}{d\theta}f(x;\theta)\,dx = 1 \quad (1)$$

Now, $\frac{d}{d\theta}\log f(x;\theta) = \frac{\frac{d}{d\theta}f(x;\theta)}{f(x;\theta)}$

so $\frac{d}{d\theta}f(x;\theta) = \left(\frac{d}{d\theta}\log f(x;\theta)\right)\cdot f(x;\theta)$

Apply ~~Add~~ $\int \cdot\, dx$ to both

$$0 \overset{(1)}{=} \int \frac{d}{d\theta}f(x;\theta)\,dx = \int\left(\frac{d}{d\theta}\log f(x;\theta)\right)f(x;\theta)\,dx = E\left[\frac{d}{d\theta}\log f(X;\theta)\right]$$