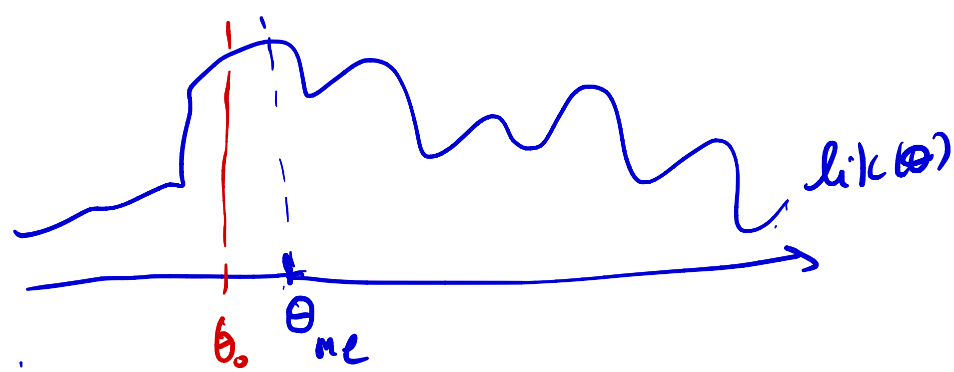
Lecture 7:

Large Sample Theory for the mle Estimator:

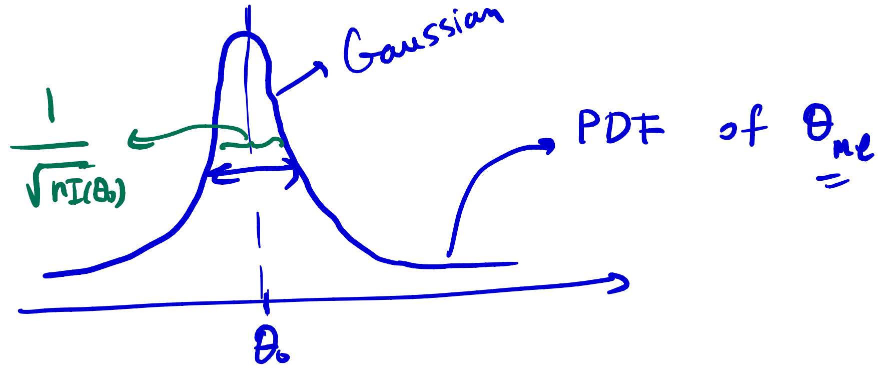
Question: How is behaviour/accuracy of the mle estimator as we increase the number of data points n?

known / unknown

Setting: $X_1, \ldots, X_n \sim f(X \mid \theta_0)$



A basic question about the mle estimator is whether/not it gives us the true parameter θ_0 as the number of data points n grows to infinity.



$X_1, \ldots, X_n \sim f(x \mid \theta_0)$

Law 1: $$\lim_{n \to \infty} \theta_{ml} = \theta_0$$

Law 2: $$\theta_{ml} \approx \mathcal{N}\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$$
$$= \theta_0 + \frac{1}{\sqrt{nI(\theta_0)}} \mathcal{N}(0,1)$$

---

when $I(\theta_0)$ is the so-called Fisher Information:

$$I(\theta_0) = E\left[\left(\frac{d}{d\theta} \log f(x \mid \theta_0) \right)^2 \right]$$

$$= E\left[\left(\frac{\frac{d}{d\theta} f(x|\theta_0)}{f(x|\theta_0)}\right)^2\right]$$

$$= \int \left(\frac{\frac{d}{d\theta} f(x|\theta_0)}{f(x|\theta_0)}\right)^2 f(x|\theta_0)\,dx$$

Implications of the laws:

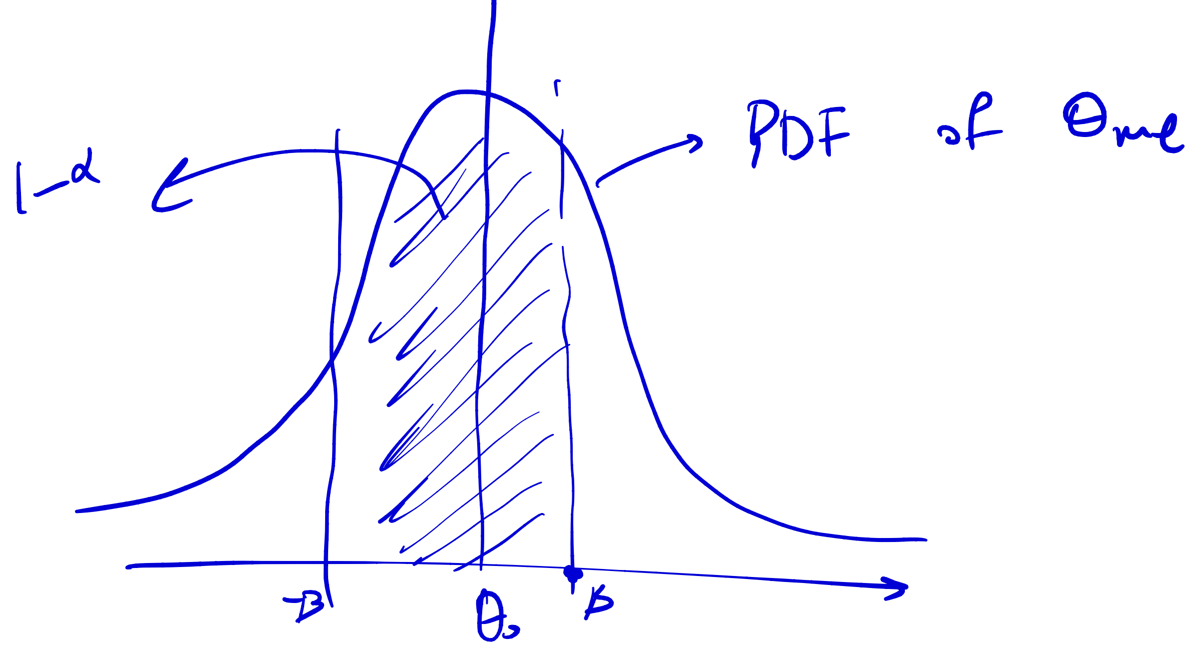
<u>Confidence intervals for θ_{me}:</u>

- We know that

$$\theta_{me} \sim \theta_0 + \frac{1}{\sqrt{n I(\theta_0)}} N(0,1)$$

- A $100(1-\alpha)\%$ confidence interval for θ_0 (from the estimator θ_{mle}) is:

$$Pr\left\{ \theta_0 \in [\theta_{mle} - \beta,\ \theta_{mle} + \beta] \right\} = 1-\alpha$$



$$\hat{\theta}_{mle} - \theta_0 \overset{\text{law 2}}{=} \frac{1}{\sqrt{nI(\theta_0)}} N(0,1)$$

$$Pr\{\theta_0 \in [\theta_{me} - \beta, \theta_{me} + \beta]\} = 1-\alpha$$

$$\Updownarrow$$

$$Pr\{\theta_{me} \in [\theta_0 - \beta, \theta_0 + \beta]\} = 1-\alpha$$

$$\Updownarrow$$

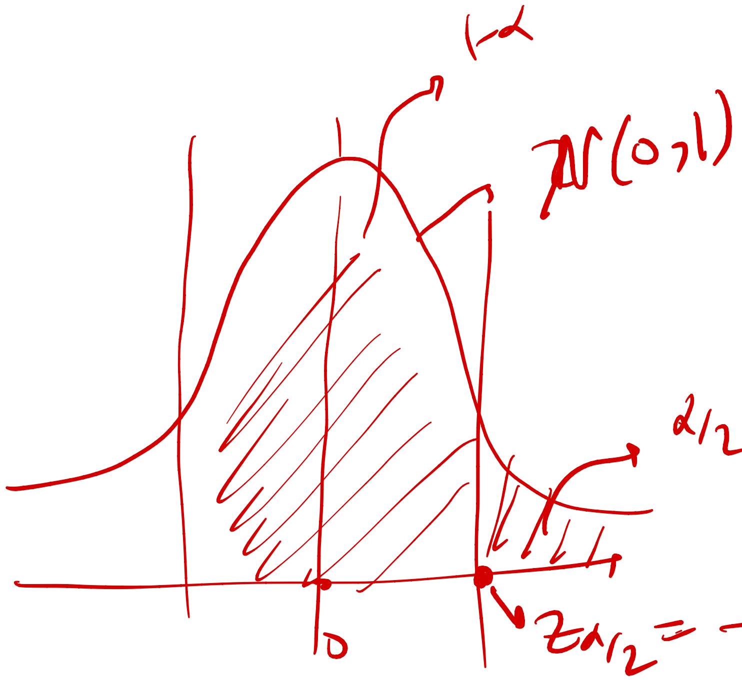
$$Pr\{\theta_0 - \beta \le \theta_{me} \le \theta_0 + \beta\} = 1-\alpha$$

$$\Updownarrow \qquad N(0,1)$$

$$Pr\left\{-\frac{\beta}{1/\sqrt{nI(\theta_0)}} \le \frac{\theta_{mle} - \theta_0}{1/\sqrt{nI(\theta_0)}} \le \frac{\beta}{\frac{1}{\sqrt{nI(\theta_0)}}}\right\} = 1-\alpha$$

$$\Updownarrow$$

$$\frac{B}{\frac{1}{\sqrt{n I(\theta_0)}}} = z_{\alpha/2} \Longrightarrow \boxed{\beta = \frac{z_{\alpha/2}}{\sqrt{n I(\theta_0)}}}$$



$$z_{\alpha/2} = \frac{B}{\frac{1}{\sqrt{n I(\theta_0)}}}$$

<u>The final confidence bound:</u>

$$\boxed{\Pr\left\{\theta_0 \in \left[\theta_{ml} - \frac{z_{\alpha/2}}{\sqrt{n I(\theta_0)}}, \theta_{ml} + \frac{z_{\alpha/2}}{\sqrt{n I(\theta_0)}}\right]\right\} = 1-\alpha}$$

The problem with the formula above is that θ_0 is not given. To resolve this issue, we can simply use θ_{me} instead of θ_0 (and the resulting error is negligible).

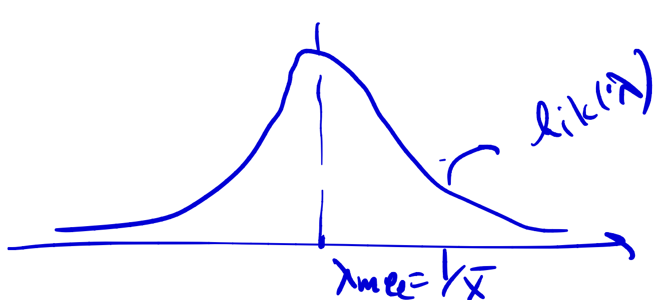
The (computable) confidence bound:

$$\Pr\left\{\theta_0 \in \left[\theta_{me} - \frac{z_{\alpha/2}}{\sqrt{n I(\theta_{me})}}, \theta_{me} + \frac{z_{\alpha/2}}{\sqrt{n I(\theta_{me})}}\right]\right\} \approx 1-\alpha$$

Example:

$X_1, \dots, X_n \sim \text{exponential}(\underline{\lambda_0})$

from last lecture:



Law 2: $\lambda_{mle} \approx \lambda_0 + \frac{1}{\sqrt{n I(\lambda_0)}} N(0,1)$

$I(\lambda_0) = ?$

$X \sim \text{exponential}(\lambda) \Rightarrow f(x|\lambda) = \lambda e^{-\lambda x} \mathbb{1}\{x \geq 0\}$

$\log(x|\lambda) = \log \lambda - \lambda x$

$\frac{d}{d\lambda} \log(x|\lambda) = \frac{1}{\lambda} - x$

$$I(\lambda_0) = \mathop{E}_{X \sim f(x|\lambda_0)}\left[\left(\frac{1}{\lambda_0} - X\right)^2\right]$$

$$= \int_{x=0}^{\infty} \left(\frac{1}{\lambda_0} - x\right)^2 \lambda_0 e^{-\lambda_0 x} \, dx$$

$$= \frac{1}{\lambda_0^2}$$

$$\lambda_{me} \approx \lambda_0 + \frac{1}{\sqrt{n \cdot \frac{1}{\lambda_0^2}}} N(0,1)$$

$$Pr\left\{ \lambda_0 \in \left[\lambda_{me} - \frac{z_{\alpha/2}}{\sqrt{n\, \underbrace{I(\lambda_0)}_{\frac{1}{\lambda_0^2} = \frac{1}{\lambda_{me}^2} = \frac{1}{\left(\frac{1}{\bar{X}}\right)^2} = (\bar{X})^2}}},\ \lambda_{me} + \frac{z_{\alpha/2}}{\sqrt{n I(\lambda_0)}} \right] \right\} = 1-\alpha$$

$$\lambda_{me} = \frac{1}{\bar{X}}$$

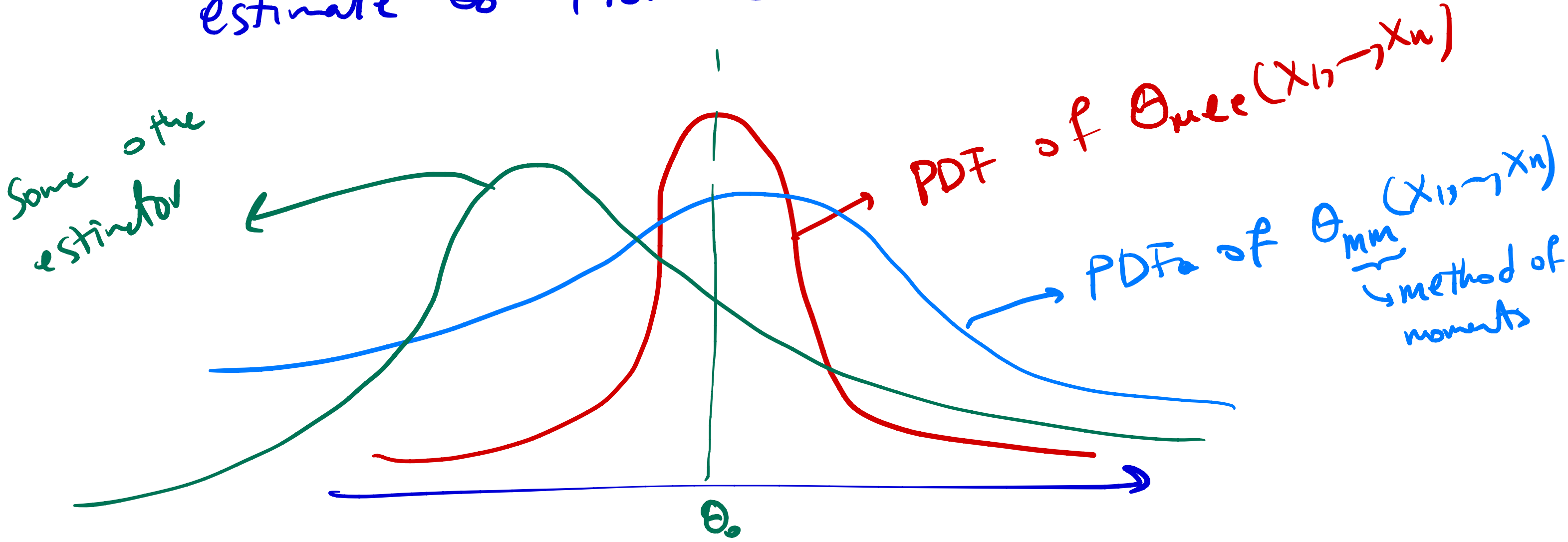
$$Pr\left\{ \lambda_0 \in \left[\frac{1}{\bar{X}} - \frac{z_{\alpha/2}}{\bar{X}\sqrt{n}},\ \frac{1}{\bar{X}} + \frac{z_{\alpha/2}}{\bar{X}\sqrt{n}} \right] \right\} = 1-\alpha$$

$$Pr\left\{ \lambda_0 \in \left[\frac{1}{\bar{X}}\left(1 - \frac{z_{\alpha/2}}{\sqrt{n}}\right),\ \frac{1}{\bar{X}}\left(1 + \frac{z_{\alpha/2}}{\sqrt{n}}\right) \right] \right\} = 1-\alpha$$

The Cramer – Rao Bound :

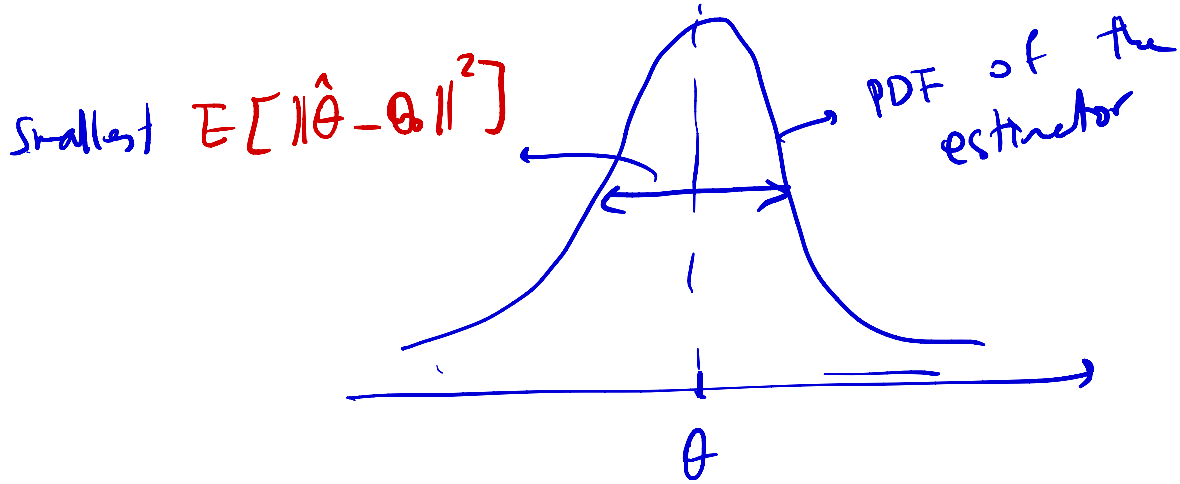
$X_1, \dots, X_n \sim f(x|\theta_0)$

estimate θ_0 from data



Recall that for our estimators, we typically prefer the one which has a smaller variance and smaller bias (or unbiased). In this regard, a fundamental question is

"what is the minimum possible varaince in terms of the underlying distribution $f(x|\theta)$ and the number of samples n?"



<u>The Cramer-Rao bound</u>: For any estimator of θ_0, call it $\hat{\theta}(X_1, \dots, X_n)$, which is <u>unbiased</u>, the variance is bounded by

$$\mathrm{Var}(\hat{\theta}) = E[\|\hat{\theta}-\theta\|^2] \geq \frac{1}{n I(\theta_0)}$$

(Law 2) Recall that for the mle estimator we have:

$$\theta_{mle} - \theta_0 \sim N\left(0, \frac{1}{nI(\theta_0)}\right)$$

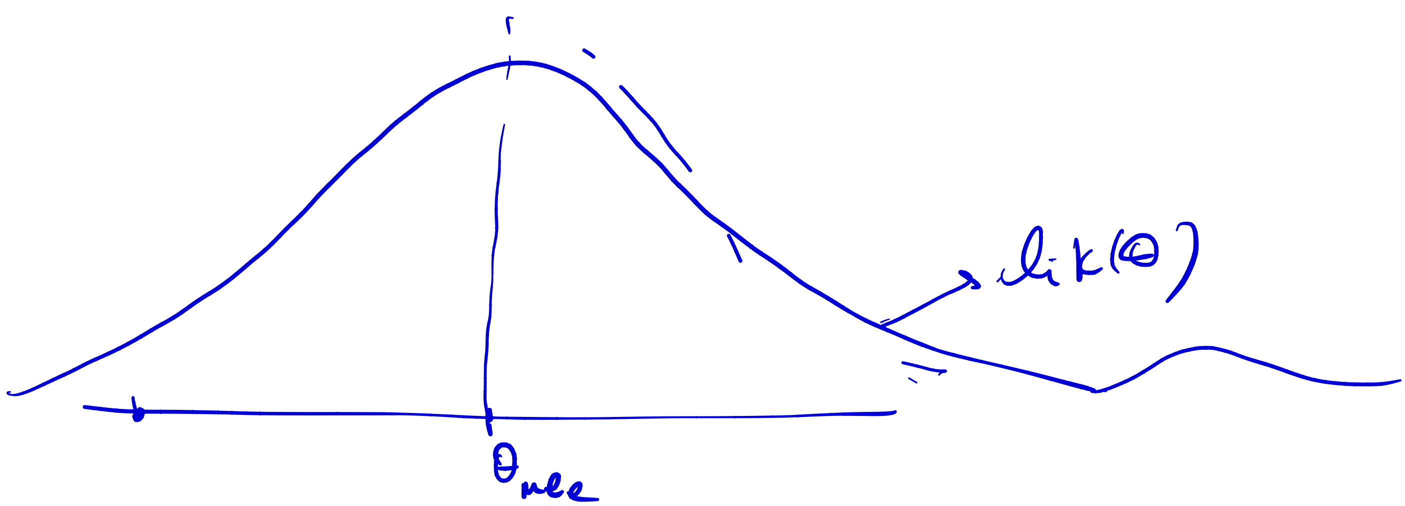
$$\Rightarrow E[\|\theta_{mle} - \theta_0\|^2] \approx \frac{1}{nI(\theta_0)}$$

Asymptotically, the variance of the mle estimator is $\dfrac{1}{n I(\theta_0)}$ which matches the minimum variance the Cramer-Rao Bound.

Proof of the Laws 1,2:

The exact proof of Laws 1,2 is beyond the scope of our class. Here, we'll provide ~~an~~ an intuitive picture/reasoning.

Law 1: $\quad \theta_{mle} \xrightarrow{n \to \infty} \theta_0$



$$\theta_{mle} = \underset{\theta}{\operatorname{argmax}}\ lik(\theta)$$

$$= \underset{\theta}{\operatorname{argmax}}\ \underbrace{\log(lik(\theta))}_{\triangleq \underline{l(\theta)}}$$

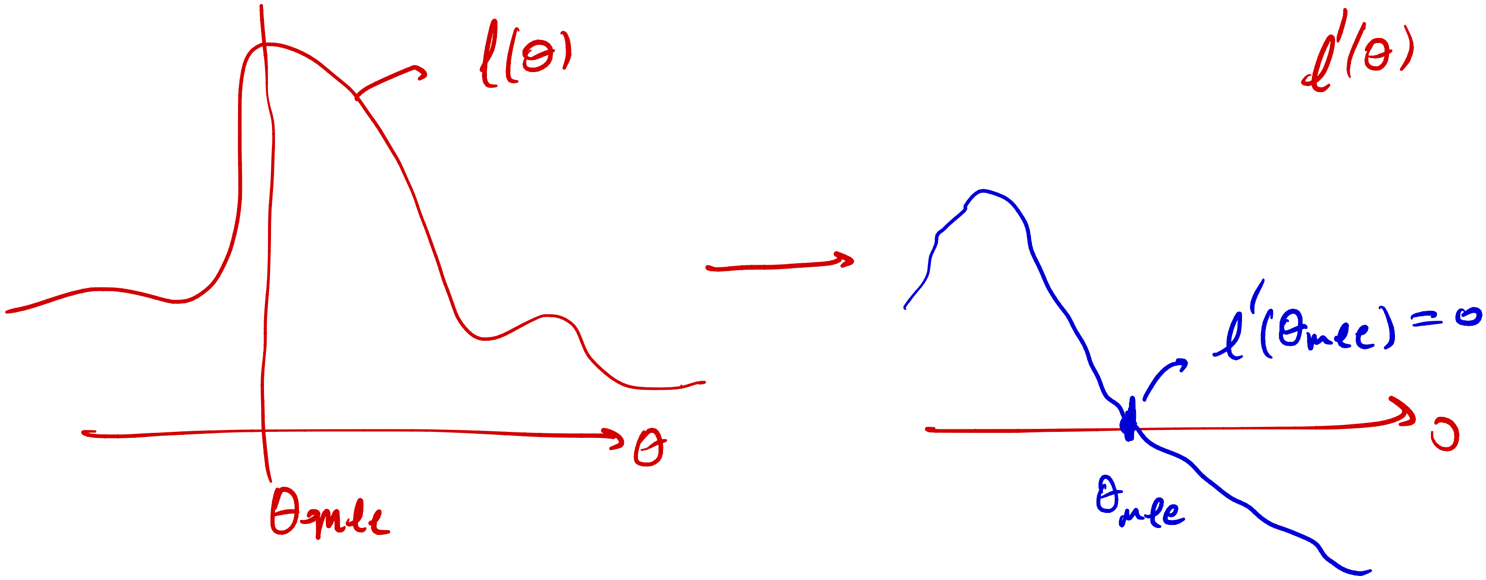
$$= \underset{\theta}{\operatorname{argmax}}\ \underbrace{l(\theta)}_{= \sum_{i=1}^{n} \log(f(x_i|\theta))}$$

We would like to show that $\theta_{mle} \to \theta_0$

$$\theta_{mle} \xrightarrow{\text{is the solution of}} l'(\theta) = 0$$

what we're going to show is that

$$l'(\underline{\theta_0}) \longrightarrow 0$$



if we show that

$\ell'(\theta_0) \approx 0$