Lecture 8:

$$\ell(\theta) = \log\left(\operatorname{lik}(\theta) \right)$$

$$= \log\left(f(x_1, \ldots, x_n \mid \theta) \right)$$

$$= \log\left(\prod_{i=1}^{n} f(x_i \mid \theta) \right)$$

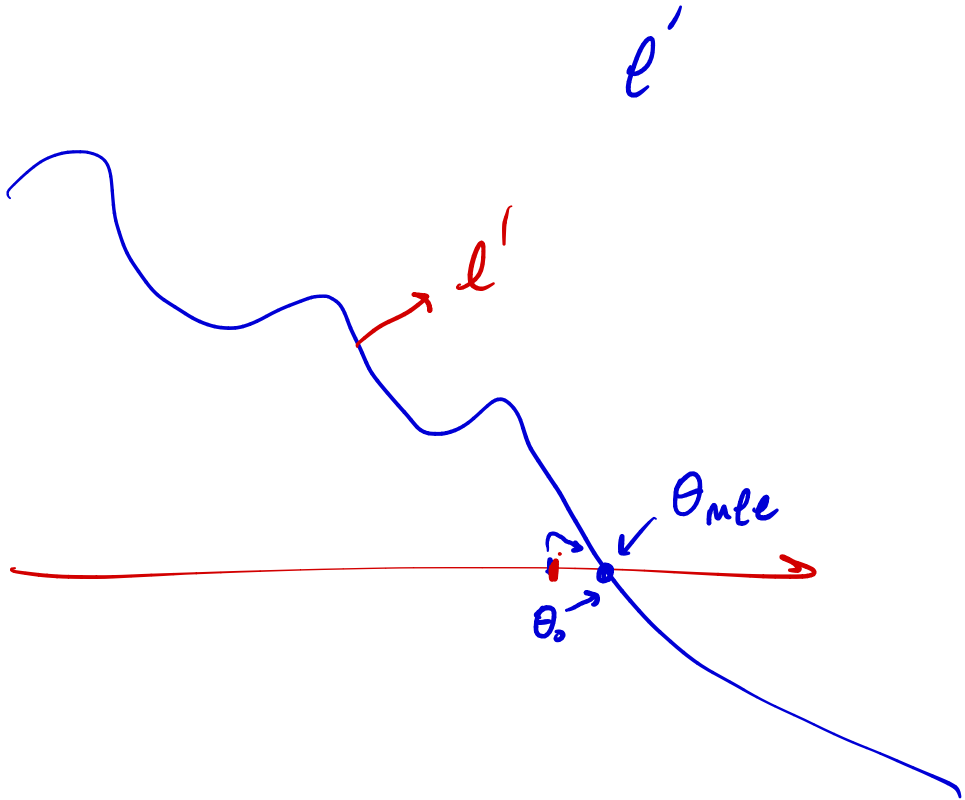
$$= \sum_{i=1}^{n} \log f(x_i \mid \theta)$$

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

Proof:

we'll show that

$$\boxed{\frac{1}{n} \underbrace{\ell'(\theta_0)}_{} \xrightarrow{n \to \infty} 0}$$.



Let's try to analyze the term

$$\frac{1}{n}\,\ell'(\theta)$$

(remember $\ell(\theta) = \sum_{i=1}^{n} \underbrace{\log f(x_i|\theta)}$)

- $$\boxed{\frac{1}{n}\ell'(\theta) = \frac{1}{n}\sum_{i=1}^{n} \underbrace{\frac{d}{d\theta}\log f(x_i|\theta)}_{Y_i \,=\, \text{function of } X_i \; g(X_i)}} \quad (*)$$

- X_i's are iid $\Longrightarrow$ Y_i's are also iid

(if Z_1 and Z_2 are independent then for any function g, $g(Z_1)$ and $g(Z_2)$ are ~~in~~ also independent from each other)

- if $\underset{\uparrow}{\overset{\text{coin} \atop \downarrow}{Z_1}} \perp\!\!\!\perp \underset{\uparrow}{\overset{\text{dice} \atop \downarrow}{Z_2}}$

then $g(Z_1) \perp\!\!\!\perp g(Z_2)$

- if Z_1 and Z_2 have the same distribution

then $g(Z_1)$ and $g(Z_2)$ also have the same distribution

(Continue from (*))

$$\frac{1}{n} \ell'(\theta) = \frac{1}{n} \underbrace{\sum_{i=1}^{n} \frac{d}{d\theta} \Big(\log f(\overset{\text{data} \sim f(x|\theta_0)}{\overset{\uparrow}{X_i}} | \theta) \Big)}_{\text{sum of } n \text{ iid random variables}}$$

$=$

$\xrightarrow[\text{CLT}]{\text{limit.}}$ behaves like a gaussian random variable

$$\frac{1}{n} \ell'(\theta) \sim N\left(\text{mean}(\theta), \frac{\text{var}(\theta)}{n}\right)$$

$$\sim \text{mean}(\theta) + \sqrt{\frac{\text{var}(\theta)}{n}}\, N(0,1)$$

$$\text{mean}(\theta) = E\left[\frac{d}{d\theta}\left(\log f(x_i \mid \theta)\right)\right]$$

$$\text{var}(\theta) = E\left[\left(\frac{d}{d\theta} \log f(x_i \mid \theta)\right)^2\right] - \left(\text{mean}(\theta)\right)^2$$

$$\frac{1}{n} \sum Y_i \xrightarrow[\text{CLT}]{} N\left(E[Y_i], \frac{\text{Var}(Y_i)}{n}\right)$$

- Let's compute mean(θ) and Var(θ) when $\theta = \theta_0$.

$$\frac{1}{n} \ell'(\theta_0) \sim N\left(\text{mean}(\theta_0), \frac{\text{Var}(\theta_0)}{n}\right)$$

$$= \text{mean}(\theta_0) + \sqrt{\frac{\text{Var}(\theta_0)}{n}}\, N(0,1)$$

$$\text{mean}(\theta_0) = E\left[\frac{d}{d\theta} \left(\log f(x|\theta_0)\right) \right]$$

$$\left((\ln u)' = \frac{u'}{u} \right)$$

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

$$\left(\text{recall that } X \sim f(x|\theta) \right)$$

$$= \int_x \frac{\frac{d}{d\theta} f(x|\theta_0)}{f(x|\theta_0)} \cdot f(x|\theta_0)\, dx$$

$$= \int_x \frac{d}{d\theta} f(x|\theta_0)\, dx$$

$$\left(\int_x \mathrm{pdf}(x)\, dx = 1 \right)$$

$$= \frac{d}{d\theta} \left(\underbrace{\int_x f(x|\theta_0)\, dx}_{=1} \right)$$

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

$$\Rightarrow \quad \frac{1}{n}\,\ell'(\theta_0) \sim \left(\sqrt{\frac{\mathrm{var}(\theta_0)}{n}}\right) \underbrace{N(0,1)}_{\text{constant}}$$

$$\underbrace{\frac{1}{n}\,\ell'(\theta_0)}_{\to 0} = \frac{\underline{\mathrm{var}(\theta_0)}\ \text{constant}}{\underbrace{\sqrt{n}}} \xrightarrow{n\to\infty} 0$$

$$\Rightarrow \quad \frac{1}{n}\,\ell'(\theta_0) \xrightarrow{n\to\infty} 0$$

$$\mathrm{mean}(\theta) = \underset{\substack{X \\ =}}{E}\left[\frac{d}{d\theta}\underbrace{\left(\log f(x|\theta)\right)}\right]$$

data $\sim f(x|\theta_0)$

$$= \int_X \frac{\frac{d}{d\theta} f(x|\theta)}{\underbrace{f(x|\theta)}} \underbrace{f(x|\theta_0)}\,dx$$

they're equal only if $\theta = \theta_0$

$$\mathrm{Var}(\theta_0) = \underbrace{E\left[\left(\frac{d}{d\theta}\log f(x|\theta_0)\right)^2\right]}_{I(\theta_0)} - \underbrace{\cancel{\mathrm{mean}}(\theta_0)^2}_{0}$$

$$= I(\theta_0) \quad (\text{fisher information})$$

$$\Longrightarrow \boxed{\underline{\frac{1}{n}\,\ell'(\theta_0)} = \sqrt{\frac{I(\theta_0)}{n}}\, N(0,1) \qquad (**)}$$

Let's now prove the second law 2.

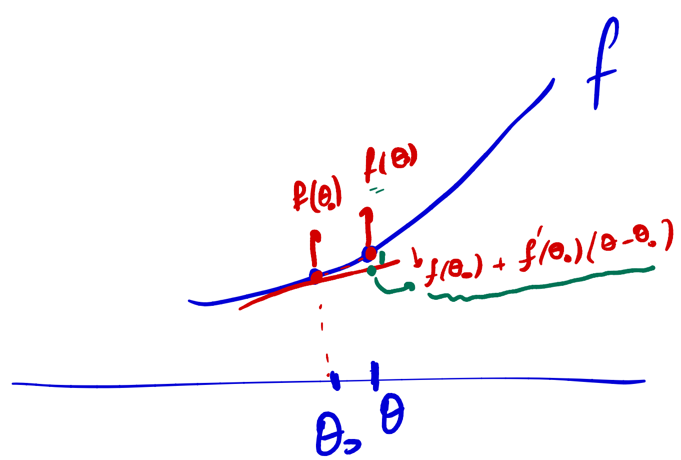
Law 2: $\theta_{ml} \sim \theta_0 + \frac{1}{\sqrt{n I(\theta_0)}} N(0,1)$

Proof (intuitive reason/picture):

$$\ell'(\theta_{me}) = 0$$

$\Rightarrow$ (let's use the Taylor approximation of the function ℓ' around the point $\theta = \theta_0$)

$$f(\theta) \approx f(\theta_0) + f'(\theta_0)(\theta - \theta_0) + \cdots$$



$$\underbrace{\ell'(\theta_{mle})}_{0} \overset{\text{taylor}}{\approx} \ell'(\theta_0) + \ell''(\theta_0)(\theta_{mle} - \theta_0) \cdots$$

$$\Rightarrow \quad \theta_{mle} - \theta_0 = - \frac{\ell'(\theta_0)}{\ell''(\theta_0)}$$

$$\Updownarrow$$

$$\theta_{mle} - \theta_0 = - \frac{\frac{1}{n}\ell'(\theta_0)}{\frac{1}{n}\ell''(\theta_0)}$$

$$\overset{(**)}{=} - \frac{\sqrt{\frac{I(\theta_0)}{\sqrt{n}}}\, N(0,1)}{\frac{1}{n}\ell''(\theta_0)} \quad (***)$$

(remember :

Law2: $\theta_{mle} - \theta_0 \sim \frac{1}{\sqrt{n I(\theta_0)}} N(0,1)$)

- The only thing left to prove is that $\frac{1}{n}\ell''(\theta_0) \xrightarrow{n \to \infty} -I(\theta_0)$ (****)

by plugging (∗∗∗∗) into (∗∗∗) we obtain

Law 2.

$$(\ast\ast\ast) = \theta_{mle} - \theta_0 = \frac{\sqrt{\frac{I(\theta_0)}{n}}\, N(0,1)}{\underbrace{-\frac{1}{n} \ell''(\theta_0)}_{(\ast\ast\ast\ast) = I(\theta_0)}}$$

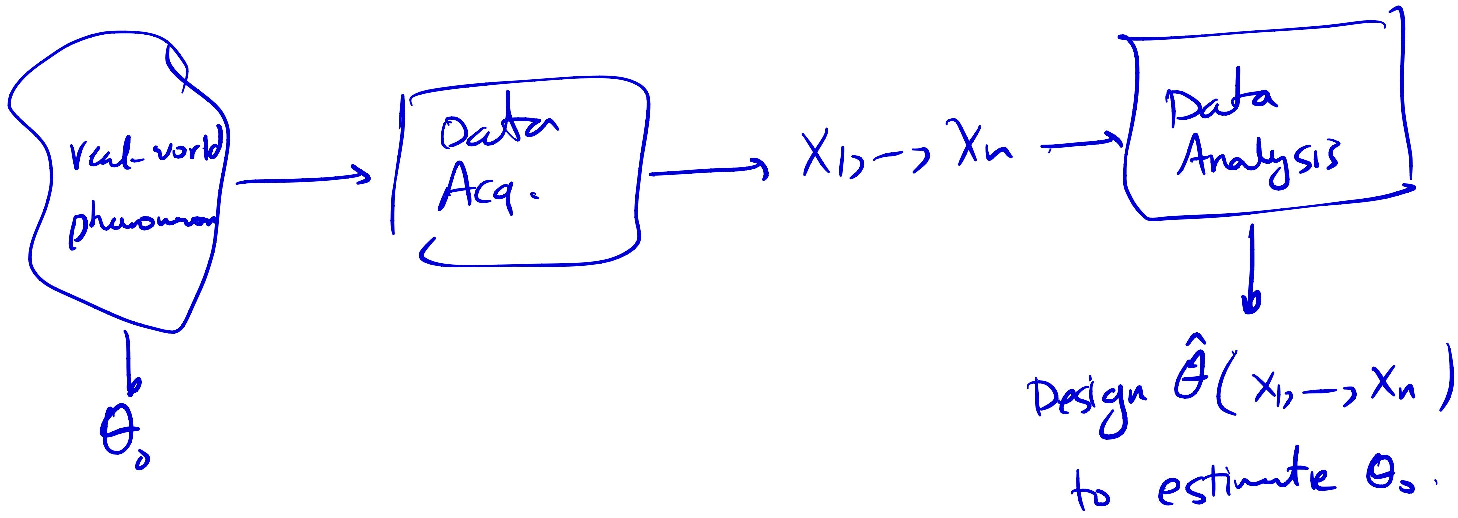
$$= \frac{1}{\sqrt{n I(\theta_0)}} N(0,1)$$

$$\frac{1}{n} \ell''(\theta_0) = \frac{1}{n}\left(\sum_{i=1}^{n} \log (f(X_i | \theta_0) \right)''$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{d}{d\theta^2} \left(\log f(X_i | \theta_0) \right)$$ → a sum of iid random variables

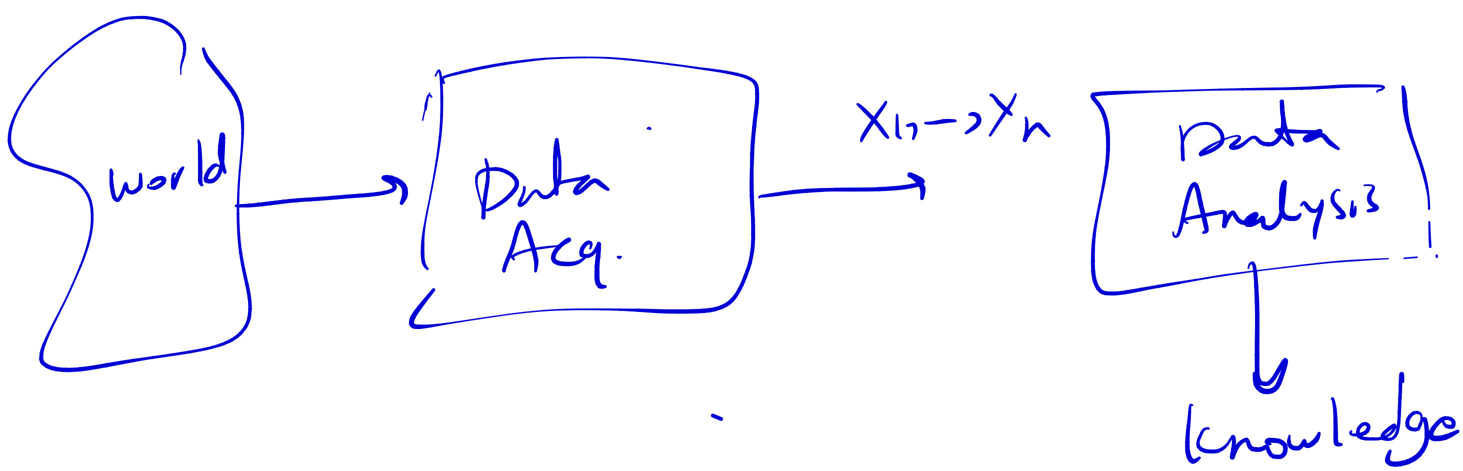
$$\xrightarrow[\text{LLN}]{\text{CLT or}} E\left[\left(\frac{d}{d\theta^2} \log f(X | \theta_0) \right) \right]$$

HW2 Q2 $\overset{\frown}{=} - I(\theta_0)$



End of Module 1

Module 2: Hypothesis Testing:



In many occasions, we can "guess" some pattern of data (using prior experience, other related phenomena, physical laws, the data itself, etc), and in order to verify this "guess", we will need to "test" our "hypothesis" (i.e. the guessed pattern) using data.

Example: The rule of thumb in most engineering systems is that the effect of nois can be modeled by adding a gaussian, and zero-mean, random variable:

$$\underbrace{observation(t)}_{\text{observation at time } t} = true\text{-}signal(t) + \underbrace{N(t)}_{\text{e.g. gaussian zero-mean}}$$

There are few assumptions/guesses to be verified:

(1) noise is additive

(2) noise is gaussian & (3) noise is zero-mean