

## Lecture 8 : Fisher Information and the MLE

- Fisher wasn't the first one to use MLE,  
but he produced the most remarkable  
result regarding its inference-  
we will go over the univariate case of  
that result today.

- we start with a 1-parameter family  
of densities.

$$\mathcal{F} = \{f_{\theta}(x), \theta \in \Omega, x \in \mathcal{X}\}$$

- we'll consider just the continuous case  
for simplicity.

- Recall we defined log-likelihood func as:

$$l_x(\theta) = \log[f_{\theta}(x)]$$

→ we'll call its derivative wrt.  $\theta$  the  
Score function

$$\dot{l}_x(\theta) = \frac{\partial}{\partial \theta} \log f_{\theta}(x) = \frac{f'_{\theta}(x)}{f_{\theta}(x)}$$

Recall the Chain Rule

How higher/lower  $\hat{l}_x(\theta)$   
gets as  $\theta$  varies.

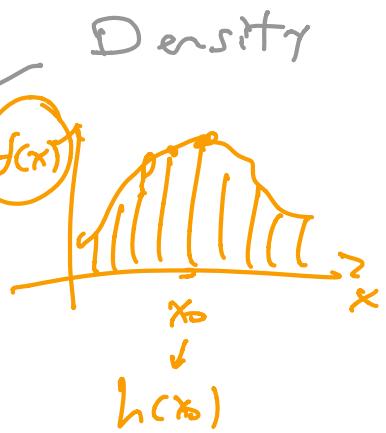
the value of the  
parameter

$$\frac{\partial}{\partial \theta} \log(f(x)) = \frac{f'(x)}{f(x)}$$

Now, let's compute the expectation of  
the score function.

$$E(x) = \int_x x f(x) dx$$

$$E[h(x)] = \int_x h(x) f(x) dx$$



$$\rightarrow E[\hat{l}_x(\theta)] = \int_x \hat{l}_x(\theta) f_\theta(x) dx$$

$$= \int_x \frac{f_\theta(x)}{f_\theta(x)} \cdot f_\theta(x) dx$$

$$= \int_x f_\theta(x) dx$$

Since density functions are continuous and  
continuously differentiable  
(the derivatives  
are popular)

$\Rightarrow$  We can differentiate inside of the integral sign.

(are continuous)

$$\text{Thus, } \int_x \frac{\partial}{\partial \theta} f_\theta(x) dx = \frac{\partial}{\partial \theta} \int f_\theta(x) dx = \frac{\partial}{\partial \theta} 1 = 0.$$

$\rightarrow$  The expectation of the score function is  $= 0$ .

The Fisher Information,  $I_\theta$  is defined as the variance of the score function.

$$V(x) = \int_x (x - E(x))^2 f(x) dx$$

$$V[h(x)] = \int_x \{h(x) - E[h(x)]\}^2 f(x) dx$$

Since  $E[\dot{l}_x(\theta)] = 0$

$$I_\theta \triangleq V[\dot{l}_x(\theta)] = \int_x [\dot{l}_x(\theta)]^2 f_\theta(x) dx$$

Now, we will show

Theorem.  $\hat{\theta}^{\text{MLE approx}} \sim N(\theta, \frac{1}{I_\theta})$

and that is nearly unbiased estimator of  $\theta$  can do better in terms of variance.

Proof: Let  $\ddot{l}_x(\theta)$  be the second derivative of  $l_x(\theta)$  wrt.  $\theta$ .

$$\ddot{l}_x(\theta) = \frac{\partial^2}{\partial \theta^2} \log f_\theta(x) = \frac{\partial}{\partial \theta} \left[ \frac{\dot{f}_\theta(x)}{f_\theta(x)} \right]$$

Quotient differentiation rule:

$$\frac{\partial}{\partial x} \frac{g(x)}{h(x)} = \frac{g'(x)h(x) - h'(x)g(x)}{[h(x)]^2}$$

$$= \frac{\ddot{f}_\theta(x) \cdot f_\theta(x) - \dot{f}_\theta(x) \cdot \dot{f}_\theta(x)}{[f_\theta(x)]^2}$$

$$= \frac{\ddot{f}_\theta(x)}{f_\theta(x)} - \left[ \frac{\dot{f}_\theta(x)}{f_\theta(x)} \right]^2$$

Now  $\ddot{l}_x(\theta)$  has expectation

$$E[\ddot{l}_x(\theta)] = \int_x \left[ \frac{\dot{f}_\theta(x)}{f_\theta(x)} \right] \cdot f_\theta(x) dx - \int_x \left[ \frac{\dot{f}_\theta(x)}{f_\theta(x)} \right]^2 f_\theta(x) dx$$

$$= \int_{-\infty}^{\infty} f_{\theta}(x) dx - \int_{-\infty}^{\infty} [\dot{l}_x(\theta)]^2 \cdot f_{\theta}(x) dx$$

Definition of  
Fisher Infor-  
mation.

$$= \frac{\partial^2}{\partial \theta^2} [1] - \overline{I}_{\theta}$$

$$= - \overline{I}_{\theta}$$

### Way point

- Defined  $\dot{l}_x(\theta)$
- Defined Fisher Inf.
- Showed  $E[\dot{l}_x(\theta)] = 0$
- Showed  $E[\ddot{l}_x(\theta)] = - \overline{I}_{\theta}$
- Stated fundamental theorem of MLE.

Now, suppose  $\underline{x} = (x_1, \dots, x_n)$  is a sample from  $f_{\theta}(x)$ . Then,

$$\dot{l}_{\underline{x}}(\theta) = \sum_{i=1}^n \dot{l}_{x_i}(\theta)$$

Bc  $\dot{l}_{\underline{x}}(\theta)$   
is in log-space.

If  $f$  bivariate  
 $f(x)$  then  $f(\underline{x}) = \prod_{i=1}^n f(x_i)$   
and  $x_i \stackrel{iid}{\sim} f(x)$

Taking the  $\log^{(C)}$

$$\dot{l}_{\underline{x}}(\theta) = \sum_{i=1}^n \dot{l}_{x_i}(\theta)$$

$$\dot{L}_x(\theta) = \sum_{i=1}^n \dot{l}_{xi}(\theta)$$

Further,  $-\ddot{L}_x(\theta) = -\sum_{i=1}^n \ddot{l}_{xi}(\theta)$

Since  $\hat{\theta}^{\text{MLE}}$  for full sample  $x$  satisfies maximizing condition  $\dot{L}_x(\hat{\theta}^{\text{MLE}}) = 0$ , we can

get Taylor Series approximation

$$0 = \dot{L}_x(\hat{\theta}^{\text{MLE}}) \approx \dot{L}_x(\theta) + \ddot{L}_x(\theta) \cdot (\hat{\theta}^{\text{MLE}} - \theta)$$

$$\Rightarrow \theta \ddot{L}_x(\theta) \approx \dot{L}_x(\theta) + \ddot{L}_x(\theta) \cdot \hat{\theta}^{\text{MLE}}$$

$$\Rightarrow \hat{\theta}^{\text{MLE}} \approx \frac{\theta \ddot{L}_x(\theta) - \dot{L}_x(\theta)}{\ddot{L}_x(\theta)}$$

$$\hat{\theta}^{\text{MLE}} \approx \theta - \frac{\dot{L}_x(\theta)}{\ddot{L}_x(\theta)}$$

Finally, recall  $\dot{L}_x(\theta) = \sum_{i=1}^n \dot{l}_{xi}(\theta)$ , so it is a sum of random variables

$X$  is r.v.  
 $f(x)$  is r.v.  $\rightarrow$  is a r.v.

while sums of r.v's are nice, we  
 have means of random variables (why?)  $\downarrow$   
CLT

(A very basic) Central Limit Theorem:

means of iid random variables are  
 Normally distributed. as  $T_n \Rightarrow \infty$

$$\text{By CLT} \dots \frac{\bar{x}(\theta)}{n} \sim N\left(0, \frac{I_0}{n}\right)$$

Convergence in distribution

Recall that if  $V(X_i) = \sigma^2$

$$V\left(\frac{\sum_{i=1}^n x_i}{n}\right) = \frac{1}{n^2} V\left(\sum_{i=1}^n x_i\right) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}$$

When  $x_i$  are iid

Then, if it is convenient to write

$$E\left(\frac{\sum_{i=1}^n x_i}{n}\right) = 0 \quad \hat{\theta} \text{ mle} \approx \theta \rightarrow \frac{\bar{x}(\theta)/n}{\bar{x}(\theta)/n} \quad E(\text{num}) = 0 \quad E(\text{den}) \neq 0$$

We know that the Numerator  $\sim N(0, I_0/n)$

Further, as  $n \rightarrow \infty$

$$\frac{\bar{x}(\theta)}{n} \rightarrow E_S[\bar{x}(\theta)] = -\frac{1}{2}\theta$$

Sample mean of r.v. converges  
 to r.v.'s expected value

Convergence  
 in value.

as  $n \rightarrow \infty$

$$\text{Thus, } \hat{\theta}^{\text{MLE}} \sim N\left(\theta, \frac{1}{n I_0}\right)$$

*(Converges in value of  $\hat{\ell}_x(\theta)$ )*

$$V\left(\theta + \frac{\dot{\ell}_x(\theta)/n}{-\ddot{\ell}_x(\theta)/n}\right) \xrightarrow{\downarrow} V\left(\frac{\dot{\ell}_x(\theta)/n}{I_0}\right) = \frac{1}{I_0^2} V\left(\frac{\dot{\ell}_x(\theta)}{n}\right) = \frac{1}{I_0 \cdot n}$$

$$= \frac{1}{I_0 n}$$

### Waypoint #2

- Wrote  $\dot{\ell}_x(\theta) = \sum_{i=1}^n \dot{\ell}_{x_i}(\theta)$
- $\ddot{\ell}_x(\theta) = \sum_{i=1}^n \ddot{\ell}_{x_i}(\theta)$
- Used  $\dot{\ell}_x(\hat{\theta}^{\text{MLE}}) \Rightarrow$  to get Taylor Series approx of  $\dot{\ell}_x(\hat{\theta}^{\text{MLE}})$  around  $\hat{\theta}^{\text{MLE}}$ .
- Solved for  $\hat{\theta}^{\text{MLE}}$  in terms of  $\theta$ ,  $\dot{\ell}_x(\theta)$  and  $\ddot{\ell}_x(\theta)$
- Used CLT to get limits for our  $\hat{\theta}^{\text{MLE}}$  expression.
- Concluded  $\hat{\theta}^{\text{MLE}} \sim N\left(\theta, \frac{1}{n \cdot I_0}\right)$

Neat!

Example : In case of Normal sampling

$$x_i \stackrel{iid}{\sim} N(\theta, \sigma^2), \quad \sigma^2 \text{ is known}$$

we know  $f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-\frac{1}{2} \left(\frac{x-\theta}{\sigma}\right)^2} \rightarrow f(x) = \left[ \frac{1}{\sqrt{2\pi\sigma^2}} \right] \cdot e^{n/2 - \frac{1}{2} \sum_{i=1}^n \left(\frac{x_i-\theta}{\sigma}\right)^2}$

$$\ell_{\bar{x}}(\theta) = -\frac{1}{2} \cdot \sum_{i=1}^n \left(\frac{x_i-\theta}{\sigma}\right)^2 - \frac{n}{2} \log(2\pi\sigma^2)$$

$$\begin{aligned} \dot{\ell}_{\bar{x}}(\theta) &= \frac{1}{\sigma^2} \cdot \sum_{i=1}^n (x_i - \theta) \\ \ddot{\ell}_{\bar{x}}(\theta) &= \frac{-n}{\sigma^2} \rightarrow -\dot{\ell}_{\bar{x}}(\theta) = \frac{n}{\sigma^2} \end{aligned}$$

$$0 = \frac{1}{\sigma^2} \left[ \sum_{i=1}^n x_i - \sum_{i=1}^n \hat{\theta}^{\text{MLE}} \right]$$

$$n \hat{\theta}^{\text{MLE}} = \sum_{i=1}^n x_i \Rightarrow \hat{\theta}^{\text{MLE}} = \frac{\sum x_i}{n}$$

Recall ..

$$E[-\dot{\ell}_{\bar{x}}(\theta)] = I_\theta$$

Then

$$I_\theta = E\left[\frac{n}{\sigma^2}\right] = \frac{n}{\sigma^2}$$

Using Fisher's Fundamental Theorem of MLE

$$\hat{\theta}^{\text{MLE}} \approx N(\theta, \frac{1}{I_\theta})$$

Fisher Inf. of  $\underline{x}$ .

$$N(\theta, \frac{\sigma^2}{n})$$

$n \cdot I_\theta$

Fisher information  
of each  $x_i$

If  $x_i$  are iid  $I_\theta$  of  $\underline{x}$   
is  $n \cdot I_\theta$  of each  $x_i$ .