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:

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

the derivative w.r.t. θ

→ We'll call its derivative w.r.t. θ the Score function

$$\dot{\ell}_{\underline{x}}(\theta) = \frac{\partial}{\partial\theta} \log f_\theta(\underline{x}) = \frac{\dot{f}_\theta(x)}{f_\theta(x)}$$

Recall the Chain Rule

How higher/lower $\ell_{\underset{\sim}{x}}(\theta)$ gets as θ varies.

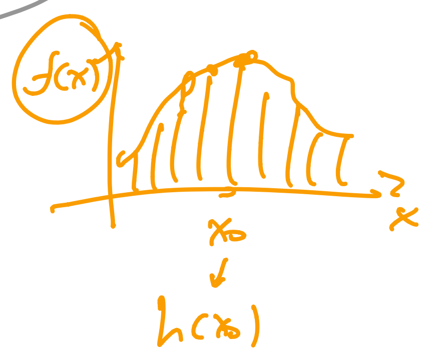
true value of the parameter

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

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

$$E(X) = \int_{\mathcal{X}} x \underbrace{f(x)}\, dx$$ Density

$$E[h(X)] = \int_{\mathcal{X}} h(x)\, f(x)\, dx$$



$$\to E[\dot{\ell}_{\underset{\sim}{x}}(\theta)] = \int_{\mathcal{X}} \dot{\ell}_x(\theta)\, f_\theta(x)\, dx$$

$$= \int_{\mathcal{X}} \frac{\dot{f}_\theta(x)}{\cancel{f_\theta(x)}} \cdot \cancel{f_\theta}(x)\, dx$$

$$= \int_{\mathcal{X}} \dot{f}_\theta(x)\, dx$$

Since popular density functions are continuous and continuously differentiable

(the derivatives

are continuous)

=> We can differentiate inside of the integral sign.

Thus, $\int_{\mathcal{X}} \frac{\partial}{\partial \theta} f_\theta(x)\,dx = \frac{\partial}{\partial \theta} \int_{\mathcal{X}} f_\theta(x)\,dx = \frac{\partial}{\partial \theta} \cdot 1$

$= 0.$

-> The expectation of the score function is $= 0$.

The Fisher Information, I_θ is defined as the variance of the score function.

$V(X) = \int_{\mathcal{X}} (X - E(X))^2 f(x)\,dx$

$V[h(x)] = \int_{\mathcal{X}} \{h(x) - E[h(x)]\}^2 f(x)\,dx$

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

$$\boxed{I_\theta \triangleq V[\dot{\ell}_x(\theta)] = \int_{\mathcal{X}} [\dot{\ell}_x(\theta)]^2 \cdot f_\theta(x)\,dx}$$

Now, we will show

Theorem. $\hat{\theta}^{MLE} \overset{aprox}{\sim} N(\theta, 1/I_\theta)$

and that no nearly-unbiased estimator of θ can do better in terms of variance.

Proof: Let $\ddot{\ell}_{\underline{x}}(\theta)$ be the second derivative of $\ell_{\underline{x}}(\theta)$ w.r.t. θ.

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

Quotient differentiation rule:

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

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

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

Now $\ddot{\ell}_{\underline{x}}(\theta)$ has expectation

$$E[\ddot{\ell}_{\underline{x}}(\theta)] = \int_{\mathcal{X}} \frac{\ddot{f}_\theta(x)}{\cancel{f_\theta(x)}} \cdot \cancel{f_\theta(x)}\, dx - \int_{\mathcal{X}} \left[\frac{\dot{f}_\theta(x)}{f_\theta(x)}\right]^2 f_\theta(x)\, dx$$

$$= \int_{\mathcal{X}} \ddot{f}_\theta(x)\,dx - \int_{\mathcal{X}} [\dot{\ell}_x(\theta)]^2 \cdot f_\theta(x)\,dx$$

Definition of Fisher Information.

$$= \frac{\partial^2}{\partial\theta^2}[1] \quad - I_\theta$$

$$= -I_\theta$$

Waypoint:

- Defined $\dot{\ell}_x(\theta)$
- Defined Fisher Inf.
- Showed $E[\dot{\ell}_{\underline{x}}(\theta)] = 0$
- Showed $E[\ddot{\ell}_{\underline{x}}(\theta)] = -I_\theta$
- Stated fundamental theorem of MLE.

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

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

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

If have $f(x)$ then $f(\underline{x}) = \prod_{i=1}^{n} f(x_i)$

and $x_i \overset{iid}{\sim} f(x)$

Taking the log (.)

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

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

Further, $-\ddot{\ell}_{\underline{x}}(\theta) = -\sum_{i=1}^{n} \ddot{\ell}_{x_i}(\theta)$

- Since $\hat{\theta}^{MLE}$ for full sample $\underline{x}$ satisfies maximizing condition $\boxed{\dot{\ell}_{\underline{x}}(\hat{\theta}^{MLE})=0}$, we can get Taylor series approximation

$$\boxed{0 = \dot{\ell}_{\underline{x}}(\hat{\theta}^{MLE}) \simeq \dot{\ell}_{\underline{x}}(\theta) + \ddot{\ell}_{\underline{x}}(\theta) \cdot (\hat{\theta}^{MLE} - \theta)}$$

$$\rightarrow \theta\, \ddot{\ell}_{\underline{x}}(\theta) \simeq \dot{\ell}_{\underline{x}}(\theta) + \ddot{\ell}_{\underline{x}}(\theta) \cdot \hat{\theta}^{MLE}$$

$$\rightarrow \hat{\theta}^{MLE} \simeq \frac{\theta\, \ddot{\ell}_{\underline{x}}(\theta) - \dot{\ell}_{\underline{x}}(\theta)}{\ddot{\ell}_{\underline{x}}(\theta)}$$

$$\boxed{\hat{\theta}^{MLE} \simeq \theta - \frac{\dot{\ell}_{\underline{x}}(\theta)}{\ddot{\ell}_{\underline{x}}(\theta)}}$$

Finally, recall $\boxed{\dot{\ell}_{\underline{x}}(\theta) = \sum_{i=1}^{n} \dot{\ell}_{x_i}(\theta)}$, so it is a $\underline{\underline{\text{sum}}}$ of $\boxed{\text{random variables}}$

X is r.v.
$f(x)$ is r.v. $\longrightarrow$ The score function is a r.v.

while sums of r.v's are nice, we love <u>means</u> of random variables (why?)

$\downarrow$

<u>CLT</u>

(A very basic) Central Limit Theorem:

Means of iid random variables are Normally distributed as $n \Rightarrow \infty$.

By CLT...
$$\frac{\dot{\ell}_{\underline{x}}(\theta)}{n} \sim N\Big(0, \underbrace{I_\theta / n}_{\text{variance}}\Big)$$

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

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

when X_i are iid

Then, it is convenient to write

$$\hat{\theta}^{MLE} \simeq \theta - \frac{\dot{\ell}_{\underline{x}}(\theta)/n}{\ddot{\ell}_{\underline{x}}(\theta)/n}$$

E (num) = 0

E (den) ≠ 0

$E\left(\frac{n}{d}\right) = 0$ as $n \to \infty$

- We know that the Numerator $\sim N(0, I_\theta/n)$
- Further, as $\underline{n \to \infty}$

$$\frac{\ddot{\ell}_{\underline{x}}(\theta)}{n} \to \boxed{E_\theta[\ddot{\ell}_{\underline{x}}(\theta)]} = -I_\theta$$

Sample mean ~~of r.v.~~ converges to r.v's expected value as $n \to \infty$

Thus, $\boxed{\hat{\theta}^{MLE} \sim N\left(\theta, \frac{1}{n\mathcal{I}_\theta}\right)}$

$$V\left(\theta - \frac{\dot{\ell}_{\underline{x}}(\theta)/n}{\ddot{\ell}_{\underline{x}}(\theta)/n}\right)$$

$$V\left(\frac{\dot{\ell}_{\underline{x}}(\theta)/n}{\ddot{\ell}_{\underline{x}}(\theta)/n}\right) \text{ as } n \to \infty \quad \underline{\mathcal{I}_\theta/n}$$

Homework

Waypoint #2

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

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

- Used $\dot{\ell}_{\underline{x}}(\hat{\theta}^{MLE}) = 0$ to get Taylor Series approx of $\dot{\ell}_{\underline{x}}(\hat{\theta}^{MLE})$ around $\hat{\theta}^{MLE}$.
- Solved for $\hat{\theta}^{MLE}$ in terms of θ, $\dot{\ell}_{\underline{x}}(\theta)$ and $\ddot{\ell}_{\underline{x}}(\theta)$
- Used CLT to get limits for our $\hat{\theta}^{MLE}$ expression.

$\boxed{\text{Neat!}}$

- Concluded $\boxed{\hat{\theta}^{MLE} \sim N\left(\theta, \frac{1}{n \cdot \mathcal{I}_\theta}\right)}$

Example: In case of Normal sampling

$X_i \overset{iid}{\sim} N(\theta, \sigma^2)$, σ^2 is known

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

we know

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

$$\dot{\ell}_{\underline{x}}(\theta) = \frac{1}{\sigma^2} \cdot \sum_{i=1}^{n}(x_i - \theta)$$

$$\ddot{\ell}_{\underline{x}}(\theta) = \frac{-n}{\sigma^2} \rightarrow \boxed{-\ddot{\ell}_{\underline{x}}(\theta) = \frac{n}{\sigma^2}}$$

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

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

Recall

$$E\left[-\ddot{\ell}_{\underline{x}}(\theta)\right] = \mathcal{I}_\theta$$

Then $\boxed{\mathcal{I}_\theta = E\left[\frac{n}{\sigma^2}\right] = \frac{n}{\sigma^2}}$

Using ~~Fisher's Fundamental~~ Theorem of MLE

$$\hat{\theta}^{MLE} \overset{apox}{\sim} N\left(\theta, \frac{1}{\mathcal{I}_\theta}\right)$$

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

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

$n \cdot I_\theta$

Fisher information of each X_i

If X_i are iid I_θ of $\underline{X}$ is $n \cdot I_\theta$ of each X_i.