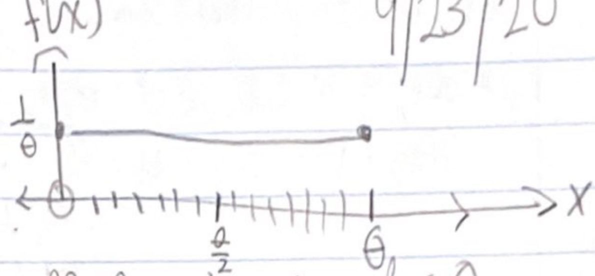
Victoria Lombardi

(1)

Math 369 9/23/20

Lecture 8

DGP $\overset{iid}{\sim} U(0,\theta)$



We want to find the MM estimator for θ

$M_1 = E[X] = \frac{0+\theta}{2} = \frac{\theta}{2} = \alpha_1(\theta) \Rightarrow \theta = 2m = \gamma(m_1)$

$\Rightarrow \hat{\theta}^{mm} = 2\hat{m}_1 = 2\bar{X}$

Data $\vec{x} = \langle 1,2,3,10 \rangle$, $\hat{\theta}^{mm} = 2\bar{x} = 2(4) = 8$

This is an absurd estimate, we're saying the true population maximum is 8 but we already seen $x_4 = 10 > 8$!! So this is clearly nonsensical

---

Another method for finding estimates/estimators goes back to the 1800's but was popularized by Fisher between 1912-1922 and it called "maximum likelihood".

$X_1, \dots, X_n \overset{iid}{\sim} DGP(\theta_1, \dots, \theta_K) = p(x; \theta_1, \dots, \theta_K)$ if discrete
$= f(x; \theta_1, \dots, \theta_K)$ if continuous

"Probability perspective" Due to independence & identical distributedness:

"density" $f(\underbrace{x_1, \dots, x_n}_{\text{inputs/variables}}; \underbrace{\theta_1, \dots, \theta_K}_{\text{givens}}) \overset{iid}{=} \prod_{i=1}^{n} f(x_i; \theta_1, \dots, \theta_K)$

"likelihood" $= \mathcal{L}(\underbrace{\theta_1, \dots, \theta_K}_{\text{inputs/variables}}; \underbrace{x_1, \dots, x_n}_{\text{givens}})$ "statistical perspective"

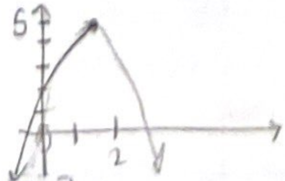
$\mathcal{L} \Leftarrow$ $= \prod_{i=1}^{n} \mathcal{L}(\theta_1, \dots, \theta_K; x_i)$ NOTE: $f > 0$ (density) $\Rightarrow \mathcal{L} > 0$

We now vary $\theta_1, \dots, \theta_K$ and try to find the values that maximize the likelihood ($\mathcal{L}$) and those values of the θ's are called the "maximum likelihood estimate(s)" (MLE).

$\hat{\theta}_1^{mle}, \dots, \hat{\theta}_K^{mle} := \underset{\Theta}{\operatorname{argmax}}(\mathcal{L}) = \operatorname{argmax}\left\{\prod_{i=1}^{n} \mathcal{L}(\theta_1, \dots, \theta_K; x_i)\right\}$

The "argmax" operator computes the argument that creates the maximum value of the function e.g.

$f(x) = x^2 + 4x - 1 = (x-2)^2 + 5$



②

$\max\{f(x)\} = 5$, $\operatorname{argmax}\{f(x)\} := \{x : f(x) = \max\{f(x)\}\} = 2$

How to find argmax. Take $f'(x) =$ set 0. And then ensure the second derivative at that value is negative.

argmax

$f'(x) = 2x + 4 \overset{set}{=} 0 \Rightarrow x_* = 2$

$f''(x) = -2$, $f''(2) = -2 < 0$ ✓

The argmax is unaffected by taking a strictly increasing function g of the set being analyzed i.e.

$$\operatorname{argmax}\{f(x)\} = \operatorname{argmax}\{g(f(x))\}$$

$$\frac{d}{dx}[g(f(x))] = g'(f(x))f'(x) \overset{set}{=} 0 \Rightarrow f'(x) = 0 \Rightarrow x_*$$

NOTE that $g(x) = \ln(x)$ is a strictly increasing function for $x > 0$

$$\hat{\theta}_1^{MLE}, \ldots, \hat{\theta}_k^{MLE} = \operatorname{argmax}\{\ln(\mathcal{L})\}$$

$$\overset{iid}{=} \operatorname{argmax}\{\ln(\prod_{i=1}^{n} f(\theta_1, \ldots \theta_k; x_i))\}$$

$l := \ln(\mathcal{L})$

$$= \operatorname{argmax}\{\sum_{i=1}^{n} \ln(f(\theta_1, \ldots, \theta_k; x_i))\}$$

$$= \operatorname{argmax}\{\sum_{i=1}^{n} l(\theta_1, \ldots \theta_k, x_i)\}$$

Why do this whole natural log thing? Well... because we're going to take the derivative of the expression inside the argmax to find the argmax & taking derivatives of sums is easy because the derivative operator is linear.

To get the MLE's, we solved the following system of equations:

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_1}[l(\theta_1, \ldots \theta_k; x_i)] \overset{set}{=} 0$$

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_2}[l(\theta_1, \ldots \theta_k; x_i)] \overset{set}{=} 0$$

$$\vdots$$

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_k}[l(\theta_1, \ldots, \theta_k; x_i)] \overset{set}{=} 0$$

It's also possible there is no maximum that corresponds to a critical point. So then you have to check the "edges" of the parameter space manually (3)

DGP: $X_1, \dots X_n \overset{iid}{\sim} Bern(\theta)$ Find $\hat{\theta}^{mle}$

$$\sum_{i=1}^{n} \frac{d}{d\theta}[\ell(\theta; x_i)] = \sum_{i=1}^{n} \frac{d}{d\theta}[\ln(p(x_i;\theta))]$$

$$\boxed{\ln(L(\theta, x_i)) \to \ln(p(x_i;\theta))}$$

$$= \sum_{i=1}^{n} \frac{d}{d\theta}[\ln(\theta^{x_i}(1-\theta)^{1-x_i})]$$

$$= \sum_{i=1}^{n} \frac{d}{d\theta}[x_i \ln(\theta) + (1-x_i)\ln(1-\theta)] = \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta}$$

$$= \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} \overset{set}{=} 0$$

$$\Rightarrow \frac{\sum x_i}{\theta} = \frac{n - \sum x_i}{1-\theta} \Rightarrow (1-\theta)\sum x_i = \theta(n - \sum x_i)$$

$$\Rightarrow \sum x_i - \theta \sum x_i = \theta n - \theta \sum x_i \Rightarrow \hat{\theta}^{mle} = \bar{X}$$

DGP: $X_1, \dots, X_n \sim N(\theta_1, \theta_2)$ Find mle's for θ_1 & θ_2

$$\sum_{i=1}^{n} \frac{\partial}{\partial\theta_1}[\underbrace{\ell(\theta_1, \theta_2; x_i)}_{=\ln(f)}] = \sum \frac{\partial}{\partial\theta_1}\left[\ln\left(\frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x_i-\theta_1)^2}\right)\right]$$

$$= \sum \frac{\partial}{\partial\theta_1}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta_2) - \frac{1}{2\theta_2}\underbrace{(x_i-\theta_1)^2}_{x_i^2 - 2x_i\theta_1 + \theta_1^2}\right]$$

$$= \sum \frac{x_i}{\theta_2} - \frac{\theta_1}{\theta_2} = \frac{\sum x_i}{\theta_2} - \frac{n\theta_1}{\theta_2} \overset{set}{=} 0 \Rightarrow \hat{\theta}^{mle} = \bar{X}$$

Now for $\hat{\theta}_2^{mle}$ ....

$$\sum \frac{\partial}{\partial\theta_2}\left[-\frac{1}{2}\ln(2\pi) - \frac{1}{2}\ln(\theta_2) - (x_i-\theta_1)^2/2\theta_2\right]$$

$$= \sum -\frac{1}{2\theta_2} + \frac{(x_i-\theta_1)^2}{2\theta_2^2} = \frac{-n}{2\theta_2} + \frac{\sum(x_i-\theta_1)^2}{2\theta_2^2} \overset{set}{=} 0$$

$$\Rightarrow \sum(x_i-\theta_1)^2 = n\theta_2 \Rightarrow \hat{\theta}_2^{mle} = \frac{1}{n}\sum(x_i-\theta_1)^2$$

plug in $\Rightarrow$ $\hat{\theta}_2^{mle} = \frac{1}{n}\sum(x_i - \hat{\theta}_1^{mle})^2 \Rightarrow \hat{\theta}_2^{mle} = \frac{1}{n}\sum(x_i-\bar{x})^2 = \hat{\sigma}^2 \neq S^2$

$\hat{\theta}^{mle} = w(x_1, \dots, x_n) \Leftrightarrow \hat{\theta}^{mle} = w(X_1, \dots, X_n)$

maximum likelihood estimate maximum likelihood estimator

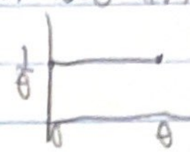
$\hat{\theta}^{mm} = w(x_1, \dots x_n) \Leftrightarrow \hat{\theta}^{mm} = w(X_1, \dots, X_n)$

DGP: $X_1, \dots, X_n \overset{iid}{\sim} U(0,\theta)$, $\hat\theta^{mm} = 2\bar{X}$, $\hat\theta^{MLE} = ?$ 4

$\sum \frac{d}{d\theta}[\ell(\theta; x_i)] = \sum \frac{d}{d\theta}\ln(f(x_i,\theta)) = \sum \frac{d}{d\theta}[\ln(\frac{1}{\theta})]$

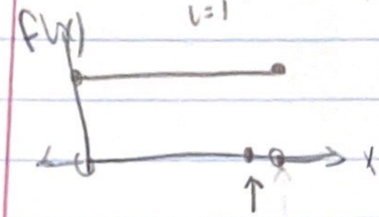
$= \sum \frac{d}{d\theta}[-\ln(\theta)] = \sum -\frac{1}{\theta} = -\frac{n}{\theta} \overset{set}{=} 0 \Rightarrow$ no solution

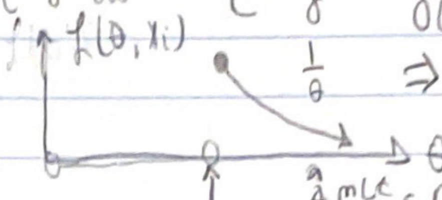
$\prod_{i=1}^{n} f(X_i;\theta) = \prod_{i=1}^{n} \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x_i \le \theta \\ 0 & \text{ow} \end{cases}$



$= \begin{cases} \frac{1}{\theta^n} & \text{if } 0 \le x_i \le \theta \ \forall x_i \\ 0 & \text{ow} \end{cases}$

$\prod_{i=1}^{n} L(\theta; x_i) = \prod_{i=1}^{n} \begin{cases} \frac{1}{\theta} & \theta > x_i \\ 0 & \text{ow} \end{cases} = \begin{cases} \frac{1}{\theta^n} & \text{if } \theta \ge x_i \ \forall x_i \\ 0 & \text{ow} \end{cases}$





$\frac{1}{\theta} \Rightarrow \hat\theta^{MLE} = \max\{X_1, \dots, X_n\}$

$\hat\theta^{MLE} = \max\{X_1, \dots, X_n\}$

We know that...

$\hat\theta^{MLE} \sim \text{Scaled Beta}(n, 1, \theta) \Rightarrow Var[\hat\theta^{MLE}] = \theta^2 \frac{n}{(n+1)(n+2)}$

$\hat\theta^{mm} = 2\bar{X} \sim ? \Rightarrow Var[2\bar{X}] = 4\frac{Var[X]}{n} = 4\frac{(\theta - 0)^2}{12n} = \theta^2 \cdot \frac{1}{3n}$

I can now compare the variance of two different estimators