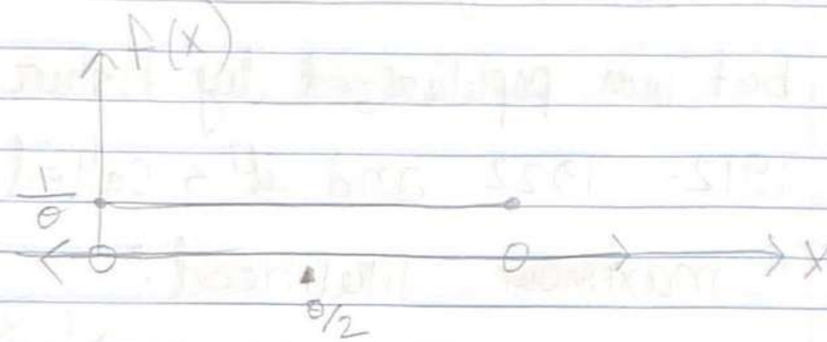
## Lecture 8:

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




We want to find the MM estimator for $\theta$.

$\mu_1 = E[X] = \frac{0+\theta}{2} = \frac{\theta}{2} = \alpha_1(\theta)$

$\Rightarrow \theta = 2\mu = \gamma_1(\mu_1)$

$\Rightarrow \hat{\theta}^{MM} = 2\hat{\mu}_1 = 2\bar{X}$.

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've 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's called : " maximum likelihood."

$X_1, \ldots, X_n \overset{iid}{\sim} DGP(\theta_1, \ldots, \theta_K)$

$\rightarrow P(X, \theta_1, \ldots \theta_K)$ if discrete

defaults

$\searrow f(X, \theta_1, \ldots \theta_K)$ if continuous

$\prod_{i=1}^{n} \mathcal{L}(\theta_1, \ldots \theta_K ; x_i) = \mathcal{L}(\theta_1, \ldots \theta_K, x_1, \ldots x_n)$

due to independence and identical distributedness

likelihood "statistic prospect"

density "probability prospect"

$\mathcal{L}(\underbrace{\theta_1, \ldots \theta_K}_{\text{input/ Variables}}, \underbrace{x_1, \ldots x_n}_{\text{givens}}) = f(\underbrace{x_1, \ldots, x_n}_{\text{inputs Variable}} ; \underbrace{\theta_1, \ldots, \theta_K}_{\text{givens}}) \overset{\text{"}}{=} \prod_{i=1}^{n} f(x_i, \theta_1, \ldots \theta_K)$

$\mathcal{L}$

Note : $f > 0$ (density) $\Rightarrow \mathcal{L} > 0$

We now vary $\theta_1, \theta_2, \ldots, \theta_K$ and try to find values that maximize the likelihood ($\mathcal{L}$) and those values of the $\theta$'s are called "maximum likelihood estimate(s)" MLE.

$$\hat{\theta}_1^{MLE}, \ldots, \hat{\theta}_K^{MLE} = \underset{\Theta}{\operatorname{argmax}}(\mathcal{L})$$

$$= \operatorname{argmax}\left\{\prod_{i=1}^{n} \mathcal{L}(\theta_1, \ldots \theta_K, x_i)\right\}$$

previous page

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

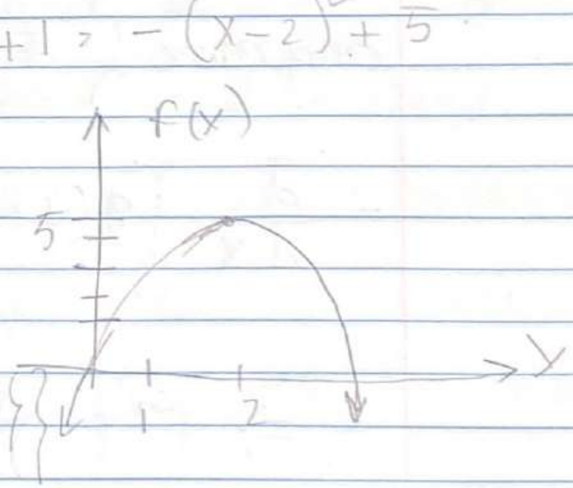
$$f(x) = -x^2 + 4x + 1 = -(x-2)^2 + 5$$

$\max\{f(x)\} = 5$

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

$\{x : f(x) = \max\{f(x)\}\}$

$= 2$



How to find argmax: Take $f'(x) =$ zero.

and then ensure the second derivative at that value is negative.

$$f'(x) = -2x + 4 \stackrel{set}{=} 0$$

$$\Rightarrow \overset{\text{argmax}}{x_*} = 2$$

$$f''(x) = -2, \quad f''(-2) = -2 < 0$$

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

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

$$= \frac{d}{dx}\left[g(f(x))\right] = \underbrace{g'(f(x))}_{>0} f'(x) \stackrel{set}{=} 0$$

$$\Rightarrow f'(x) = 0 \Rightarrow x_*$$

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

---

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

$$\overset{iid}{=} \operatorname{argmax}\left\{\ln\left(\prod_{i=1}^{n} \mathcal{L}(\theta_1, \dots \theta_K, x_i)\right)\right\}$$

$$= \operatorname{argmax}\left\{\sum_{i=1}^{n} \ln(\mathcal{L}(\theta_1, \dots \theta_K; x_i))\right\}$$

Why do this whole log thing? Well because we are going to take the derivative of the expression inside the argmax to find the arg max and taking derivatives of sums is easy because the derivative operator is linear. To get MLE'S we solved the system of equations:

$l$ : $\ln(\mathcal{L}) = \text{argmax}\left\{\sum_{i=1}^{n} l(\theta_1, \ldots, \theta_k, x_i)\right\}$

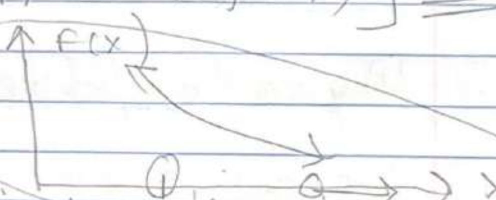
$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_1}\left[l(\theta_1, \ldots \theta_k; x_i)\right] \overset{set}{=} 0,$$

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

$\vdots$

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

add to Precal sec




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

* DGP: $X_1, \ldots X_n \overset{iid}{\sim} \text{Bern}(\theta)$. Find $\hat{\theta}^{MLE}$

$$\sum_{i=1}^{n} \frac{d}{d\theta}\left[\ell(\theta, x_i)\right]$$

$\ell n(\mathcal{L}(\theta, x_i)) \rightarrow \ln(P(X_i; \theta))$

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

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

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

$$= \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta}$$

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

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

$$\Rightarrow \frac{\Sigma x_i}{\theta} = \frac{n - \Sigma x_i}{1-\theta}$$

$$\Rightarrow (1-\theta)\Sigma x_i = \theta(n - \Sigma x_i)$$

$$\Rightarrow \Sigma x_i - \cancel{\theta \Sigma x_i} = n\theta - \cancel{\theta \Sigma x_i}$$

$$\Rightarrow \boxed{\hat{\theta} = \frac{\Sigma x_i}{n} = \bar{X}}$$ # important

☆ DQP: $X_1, \ldots X_n \sim N(\theta_1, \theta_2)$. Find MLE for $\theta_1$ and $\theta_2$.

$$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_1}\left[ l\overset{= \ln(f)}{(\theta_1, \theta_2 ; x_i)} \right]$$

$$= \Sigma \frac{\partial}{\partial \theta_1}\left[ \ln\left( \frac{1}{\sqrt{2\pi\theta_2}} e^{-\frac{1}{2\theta_2}(x_i - \theta_1)^2} \right) \right]$$

$x_i^2 - 2x_i\theta_1 + \theta$

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

$$-\frac{x_i}{2\theta_2} + \frac{x_i\theta_1}{\theta_2} - \frac{\theta_1^2}{2\theta_2}$$

$$= \sum\left(\frac{x_i}{\theta_2} - \frac{\theta_1}{\theta_2}\right)$$

$$= \frac{\sum x_i}{\theta_2} - \frac{n\theta_1}{\theta_2} \overset{set}{=} 0$$

$$\Rightarrow \hat{\theta}_1^{MLE} = \bar{X}$$

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

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

$$= \Sigma\left(-\frac{1}{2\theta_2} + \frac{(x_i-\theta_1)^2}{2\theta_2^2}\right)$$

$$= -\frac{n}{\cancel{2}\theta_2} + \frac{\Sigma(x_i-\theta_1)^2}{\cancel{2}\theta_2^2} \overset{\text{set}}{=} 0$$

$$\Rightarrow \Sigma(x_i-\theta_1)^2 = n\theta_2$$

$$\Rightarrow \hat{\theta}_2^{MLE} = \frac{1}{n}\Sigma(x_i-\theta_1)^2$$

$$\overset{\text{plugin}}{\Longrightarrow} \hat{\theta}_2^{MLE} = \frac{1}{n}\Sigma\left(x_i-\hat{\theta}_1^{MLE}\right)^2$$

$$\Rightarrow \hat{\theta}_2^{MLE} = \frac{1}{n}\Sigma(x_i-\bar{x})^2$$

$$= \hat{\sigma}^2 \neq s^2$$

$\underset{\sim}{\hat{\theta}}^{MLE} = w(x_1, \; x_n) \Leftrightarrow \underset{\sim}{\hat{\theta}}^{MLE} = w(X_1, \; X_n)$

Maximum likelihood estimate | maximum likelihood estimator.

$\hat{\theta}^{MM} = W(x_1, x_n) \Leftrightarrow \hat{\theta}^{MM} = W(X_1, \; X_n)$

DGP: $X_1, \; X_n \overset{iid}{\sim} U(0,1)$, $\hat{\theta}^{MM} = 2\bar{X}$, $\hat{\theta}^{MLE} = ?$

$$\mathcal{E} \frac{d}{d\theta}\left[\ell(\theta; x_i)\right] = \mathcal{E} \frac{d}{d\theta}\left[\ln(f(x_i;\theta))\right]$$

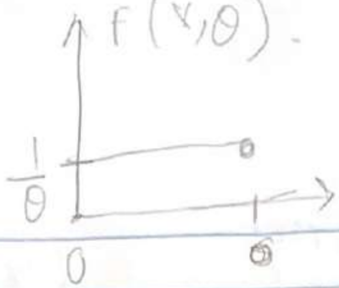
$$= \mathcal{E} \frac{d}{d\theta}\left[\ln\left(\frac{1}{\theta}\right)\right]$$

$$= \mathcal{E} \frac{d}{d\theta}\left[-\ln(\theta)\right]$$

$$= \mathcal{E} -\frac{1}{\theta} = -\frac{n}{\theta} \overset{set}{=} 0$$

$$\Rightarrow -\frac{n}{\theta} \overset{set}{=} 0 \Rightarrow \text{no sol.}$$

This is a density



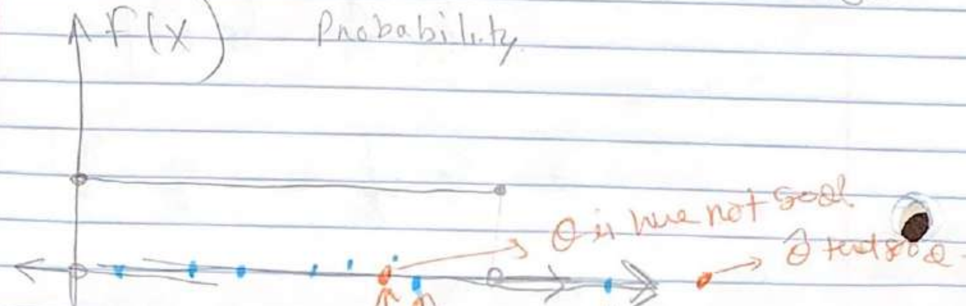

$$\prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \begin{cases} \frac{1}{\theta} & \text{if } 0 \le x_i \le \theta \\ 0 & \text{Otherwise} \end{cases}$$

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

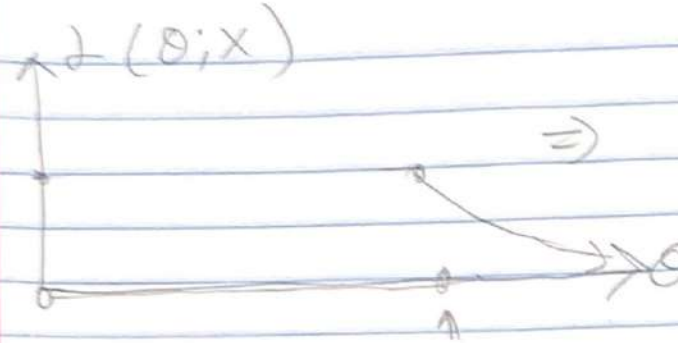
---

$$\prod_{i=1}^{n} L(\theta, x_i) = \prod_{i=1}^{n} \begin{cases} 1/\theta & \theta > x_i \\ 0 & \text{o/t} \end{cases}$$

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



Likelihood:




$\Rightarrow \hat{\theta}^{MLE} = \text{Max}\{x_1, \dots, x_n\}$

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

Beyond scope of course from 368

we know that:

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

$$= \theta^2 \frac{n}{(n+1)(n+2)}$$

$\hat{\theta}^{MM} = 2\bar{X} \sim ? \Rightarrow$

$$Var[2\bar{x}] = 4 \frac{Var[x]}{n} = \frac{4(\theta - 0)^2}{12n}$$

$$= \theta^2 \frac{1}{3n}.$$

I can now compare the variance of two different estimators.