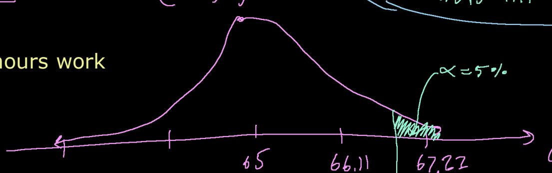
$H_a: \theta > 65$, $H_0: \theta \le 65$, $\alpha = 5\%$ DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta, 3.5^2)$

$\Rightarrow \bar{X}_n \sim N\left(\theta, \left(\frac{3.5}{\sqrt{n}}\right)^2\right)$ exact

$\bar{x}$ $\hat{\theta} | H_0 \sim N\left(65, \left(\underbrace{\frac{3.5}{\sqrt{10}}}_{1.11}\right)^2\right)$

$65 + 1.645 \cdot 1.11$

Office hours work



$\hat{\theta}$ $\left\{ \begin{matrix} REJ \\ | \\ \hat{\theta} \end{matrix} \right.$

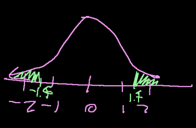
$\longleftarrow$ RET $\longrightarrow$ $\underset{\hat{\theta}_c}{66.83}$ $\longleftarrow$ Rej $\longrightarrow$

$= (-\infty, 66.83]$ $\qquad$ $= (66.83, \infty)$

CDF of $Z \sim N(0,1)$

$\Phi(x) := F_Z(x)$

$P\left(\hat{\theta} | H_0 > \hat{\theta}_c\right) = \alpha = 5\%$



$\Rightarrow P\left(\frac{\hat{\theta}|H_0 - 65}{1.11} > \frac{\hat{\theta}_c - 65}{1.11}\right) = 5\%$

$\Rightarrow P\left(Z > \underbrace{\frac{\hat{\theta}_c - 65}{1.11}}_{z}\right) = 5\%$

$P(Z > \overset{1.645}{z}) = 5\%$

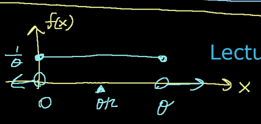
$= P(Z < -z) = \Phi(-z) = 5\%$, $-z = -1.645$

DGP: iid $Bern(\theta)$

$\bar{X} \sim N\left(\theta, \sqrt{\frac{\theta(1-\theta)}{n}}\right)$

$1.645 = \frac{\hat{\theta}_c - 65}{1.11} \Rightarrow \hat{\theta}_c = 65 + 1.645 \cdot 1.11$

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



Lecture 8 begins here

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

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".

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

Due to independence and identical distributedness,

likelihood "statistics perspective" $\quad$ density "probability perspective"

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

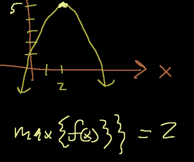
Note: $f > 0$ (density) $\Rightarrow \mathcal{L} > 0$.

We now vary theta_1, ..., theta_K and try to find the values that maximize the likelihood (curly-L) and those values of the thetas are called the "maximum likelihood estimate(s)" (MLE).

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

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

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



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

$\{x : f(x) = \max\{f(x)\}\} = 2$

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

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

$f''(x) = -2$, $f''(2) = -2 < 0$ ✓

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

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

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

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

$\hat{\hat{\theta}}_1^{MLE}, \ldots, \hat{\hat{\theta}}_K^{MLE} = \text{argmax}\{\ln(\mathcal{L})\} \overset{iid}{=} \text{argmax}\left\{\ln\left(\prod_{i=1}^{n} \mathcal{L}(\theta_1, \ldots, \theta_K; x_i)\right)\right\}$

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

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

$= \text{argmax}\left\{\sum_{i=1}^{n} \ell(\theta_1, \ldots, \theta_K; x_i)\right\}$

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 and taking derivatives of sums is easy because the derivative operator is linear.

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

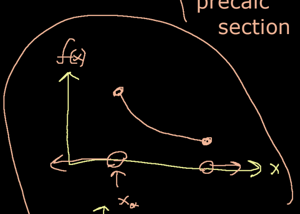
$\sum_{i=1}^{n} \frac{\partial}{\partial \theta_1}\left[\ell(\theta_1, \ldots \theta_K; x_i)\right] \overset{set}{=} 0,$

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

$\vdots$

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

add to precalc section



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

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

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

$= \sum_{i=1}^{n} \frac{d}{d\theta}\left[x_i \ln(\theta) + (1 - x_i)\ln(1-\theta)\right] = \sum_{i=1}^{n} \frac{x_i}{\theta} - \frac{1-x_i}{1-\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} \overset{set}{=} 0$

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

$\Rightarrow \sum x_i - \cancel{\theta \sum x_i} = \theta n - \cancel{\theta \sum x_i} \Rightarrow \hat{\hat{\theta}}^{MLE} = \bar{x}$

DGP: $X_1, \ldots, X_n \overset{iid}{\sim} N(\theta_1, \theta_2)$. Find MLE's for theta_1 and theta_2

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

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

$= \sum \frac{x_i}{\theta_2} - \frac{\theta_1}{\theta_2} = \frac{\sum x_i}{\cancel{\theta_2}} - \frac{n\theta_1}{\cancel{\theta_2}} \overset{set}{=} 0 \Rightarrow \hat{\hat{\theta}}_1^{MLE} = \bar{x}$.

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

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

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

$\Rightarrow \sum(x_i - \theta_1)^2 = n\theta_2 \Rightarrow \hat{\hat{\theta}}_2^{MLE} = \frac{1}{n}\sum(x_i - \theta_1)^2 \overset{\text{plug in}}{\Rightarrow} \hat{\hat{\theta}}_2^{MLE} = \frac{1}{n}\sum\left(x_i - \hat{\hat{\theta}}_1^{MLE}\right)^2$

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

$\hat{\hat{\theta}}^{MLE} = w(x_1, \ldots, x_n) \Longleftrightarrow \hat{\theta}^{MLE} = w(X_1, \ldots, X_n)$

maximum likelihood estimate $\qquad$ maximum likelihood estimator

$\hat{\hat{\theta}}^{MM} = w(x_1, \ldots, x_n) \Longleftrightarrow \hat{\theta}^{MM} = w(X_1, \ldots, X_n)$

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

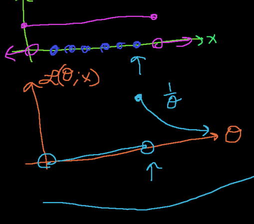
$\sum \frac{d}{d\theta}\left[\ell(\theta; x_i)\right] = \sum \frac{d}{d\theta}\left[\ln(f(x_i; \theta))\right] = \sum \frac{d}{d\theta}\left[\ln\left(\frac{1}{\theta}\right)\right] = \sum \frac{d}{d\theta}\left[-\ln(\theta)\right]$

$= \sum -\frac{1}{\theta} = -\frac{n}{\theta} \overset{set}{=} 0 \Rightarrow$ no solution

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

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

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



$\Rightarrow \hat{\hat{\theta}}^{MLE} = \max\{x_1, \ldots, x_n\}$

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

Beyond scope of course... From 368 we know that...

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

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

I can now compare the variance of two different estimators.