Maximum likelihood estimation (MLE) — 2^{nd} way of constructing estimators

$X_1, \cdots, X_n \overset{iid}{\sim}$ distribution w/ parameter θ

if the distribution is $\begin{cases} \text{continuous : pdf } f(x;\theta) \\ \text{discrete : pmf } P(X=x;\theta) \end{cases}$

e.g. $X_1, \cdots, X_n \overset{iid}{\sim} N(\theta, 1)$, $f(x;\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}}$

$X_1, \cdots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$, $P(X=x;\theta) = \frac{\theta^x \cdot e^{-\theta}}{x!}$

Joint $\begin{cases} \text{density (pdf) } f(x_1, \cdots, x_n;\theta) \overset{iid}{=} f(x_1;\theta)\cdots f(x_n;\theta) \overset{e.g.}{=} \left(\frac{1}{\sqrt{2\pi}}\right)^n e^{-\frac{\sum_{i=1}^n (x_i-\theta)^2}{2}} \\ \text{pmf } P(X_1=x_1, \cdots, X_n=x_n;\theta) \overset{iid}{=} P(X_1=x_1;\theta)\cdots P(X_n=x_n;\theta) \overset{e.g.}{=} \frac{\theta^{\sum_{i=1}^n x_i} \cdot e^{-n\theta}}{x_1! \cdots x_n!} \end{cases}$

$\underbrace{x_1, \cdots, x_n}_{\text{data}}$ are realizations of $X_1, \cdots, X_n$

$f(x_1, \cdots, x_n;\theta)$ or $P(X_1=x_1, \cdots, X_n=x_n;\theta)$: the "chance" of observing the data given parameter θ

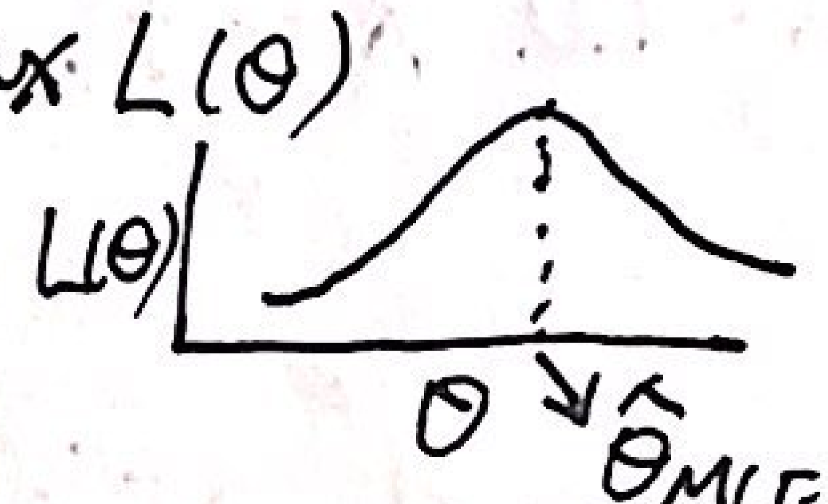
MLE: Find the value of θ such that the "chance" of observing the data is maximized.

Definition of "likelihood": the likelihood of θ is defined as

$$L(\theta) = \begin{cases} f(x_1, \cdots, x_n;\theta) & \text{if } X_1, \cdots, X_n \text{ are continuous} \\ P(X_1=x_1, \cdots, X_n=x_n;\theta) & \text{if } X_1, \cdots, X_n \text{ are discrete} \end{cases}$$

MLE of θ: $\hat{\theta}_{MLE}$ = maximizer of $L(\theta) = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$

e.g. $L(\theta) = -(\theta-1)^2 \Rightarrow \underset{\theta}{\operatorname{argmax}}\, L(\theta) = 1$



Examples:

1. $X_1, \cdots, X_n \overset{iid}{\sim} \text{Poisson}(\theta)$. $\theta > 0$ Find $\hat{\theta}_{MLE}$.

The likelihood of θ is

$$L(\theta) = P(X_1, \cdots, X_n;\theta) \overset{iid}{=} P(X_1;\theta)\cdots P(X_n;\theta) = \frac{\theta^{\sum_{i=1}^n X_i} e^{-n\theta}}{\prod_{i=1}^n X_i!}$$

log-likelihood: $\ell(\theta) = \log L(\theta)$ Fact: $\underset{\theta}{\operatorname{argmax}}\, \ell(\theta) = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$

$$\ell(\theta) = \log L(\theta) = \left(\sum_{i=1}^n X_i\right) \cdot \log\theta - n\theta - \log\left(\prod_{i=1}^n X_i!\right)$$

$$\frac{d}{d\theta}\ell(\theta) = \left(\sum_{i=1}^n X_i\right)\cdot\frac{1}{\theta} - n \qquad \text{Set } \frac{d}{d\theta}\ell(\theta) = 0 \Rightarrow \theta = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X_n}$$

$$\frac{d^2}{d\theta^2}\ell(\theta) = \underbrace{\left(\sum_{i=1}^n X_i\right)}_{\geq 0}\cdot\left(-\frac{1}{\theta^2}\right) \leq 0$$

Hence, $\overline{X_n} = \underset{\theta}{\operatorname{argmax}}\, \ell(\theta) \Rightarrow \hat{\theta}_{MLE} = \overline{X_n} = \hat{\theta}_{MOM}$.

2. $X_1, \cdots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Find $\hat{\mu}_{MLE}$, $\hat{\sigma}^2_{MLE}$.

The likelihood of $\theta = \begin{pmatrix}\mu \\ \sigma^2\end{pmatrix}$ is

$$L(\theta) = L(\mu, \sigma^2) = f(X_1, \cdots, X_n; \theta) \overset{iid}{=\!=} f(X_1;\theta)\cdots f(X_n;\theta)$$

$$= \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_1-\mu)^2}{2\sigma^2}} \cdots \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(X_n-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sqrt{2\pi}\cdot\sigma}\right)^n \cdot e^{-\frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}}$$

The log-likelihood is

$$\ell(\theta) = \ell(\mu; \sigma^2) = \log L(\mu, \sigma^2) = -n\cdot\log(\sqrt{2\pi}\sigma) - \frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}$$

$$= -n\log(\sqrt{2\pi}) - \frac{n}{2}\log(\sigma^2) - \frac{\sum_{i=1}^n (X_i-\mu)^2}{2\sigma^2}$$

Take partial derivatives;

$$\begin{cases} \frac{\partial \ell(\theta)}{\partial \mu} = -\frac{\sum_{i=1}^n 2(X_i-\mu)\cdot(-1)}{2\sigma^2} = \frac{\sum_{i=1}^n (X_i-\mu)}{\sigma^2} \overset{\text{set to}}{=\!=} 0 \Rightarrow \mu = \frac{1}{n}\sum_{i=1}^n X_i = \overline{X_n} \\ \frac{\partial \ell(\theta)}{\partial(\sigma^2)} = -\frac{n}{2}\cdot\frac{1}{\sigma^2} - \frac{\sum_{i=1}^n (X_i-\mu)^2}{2}\left(-\left(\frac{1}{\sigma^2}\right)^2\right) \overset{\text{set to}}{=\!=} 0 \end{cases}$$

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

$$\Longleftrightarrow n\left(\cancel{\frac{1}{\sigma^2}}\right) + \left(\sum_{i=1}^n (X_i-\mu)^2\right)\cdot\left(-\left(\frac{1}{\sigma^2}\right)^{1}\right) = 0$$

$$\Longleftrightarrow \sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i-\mu)^2$$

$$\Longleftrightarrow \begin{cases} \mu = \overline{X_n} \\ \sigma^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})^2 \end{cases} \Rightarrow \text{Hessian matrix} \begin{bmatrix} \frac{\partial^2 \ell(\theta)}{\partial \mu^2} & \frac{\partial^2 \ell(\theta)}{\partial\mu\cdot\partial(\sigma^2)} \\ \frac{\partial^2 \ell(\theta)}{\partial(\sigma^2)\cdot\partial\mu} & \frac{\partial^2 \ell(\theta)}{\partial(\sigma^2)^2} \end{bmatrix}$$

$\hat{\mu}_{MLE}$ is unbiased
$\hat{\sigma}^2_{MLE}$ is biased $\Leftarrow \begin{cases} \hat{\mu}_{MLE} = \overline{X_n} \\ \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^n (X_i - \overline{X_n})^2 \end{cases}$

verify it is negative definite at $\begin{pmatrix}\mu \\ \sigma^2\end{pmatrix} = \begin{pmatrix}\overline{X_n} \\ \frac{1}{n}\Sigma(X_i-\overline{X_n})^2\end{pmatrix}$

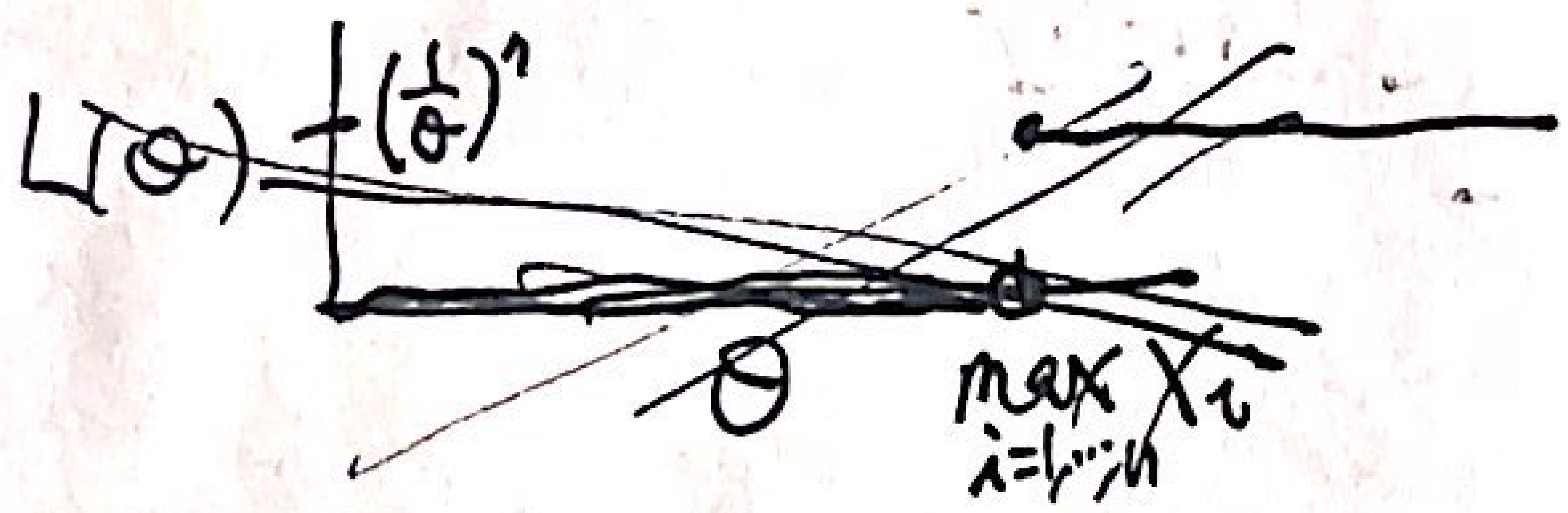
$X_1, \dots, X_n \overset{iid}{\sim} \text{Unif}(0, \theta) \quad \theta > 0$. Find $\hat{\theta}_{MLE}$

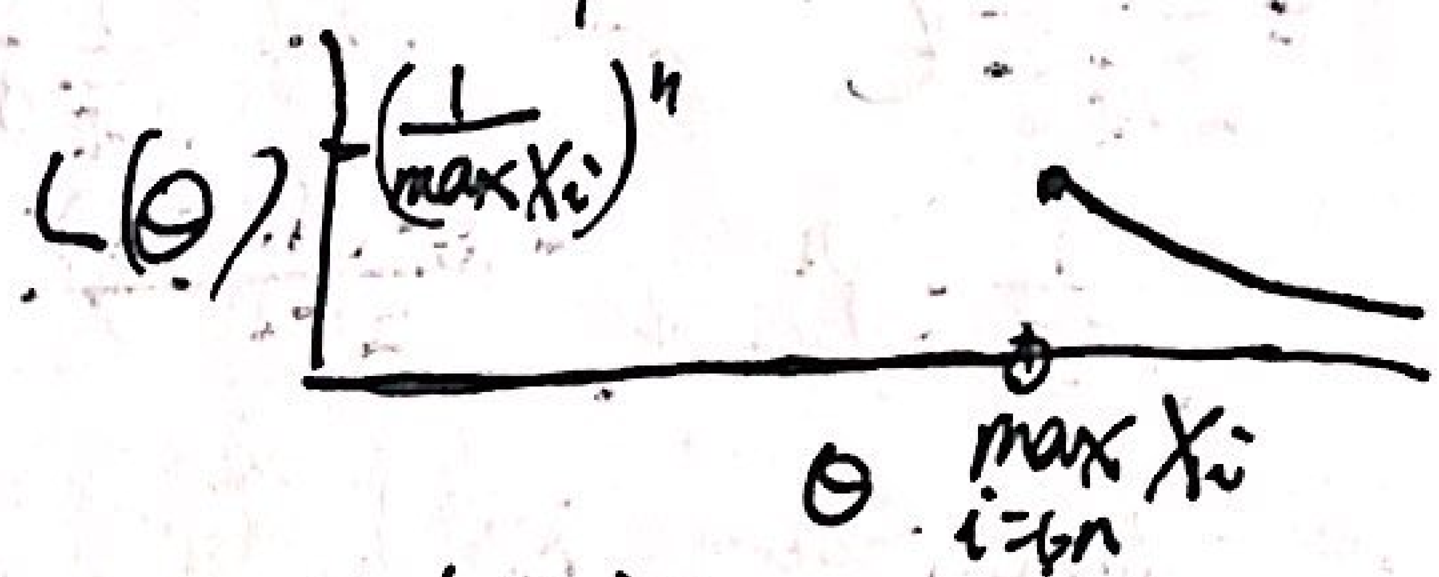
$$L(\theta) = f(X_1, \dots, X_n; \theta) \overset{iid}{=} f(X_1; \theta) \cdots f(X_n; \theta)$$

$$f(X_i; \theta) = \begin{cases} \frac{1}{\theta} & \text{if } X_i \in [0, \theta] \\ 0 & \text{otherwise} \end{cases} = \frac{1}{\theta} \cdot I(X_i \le \theta)$$

$$\begin{aligned} L(\theta) &= \frac{1}{\theta} \cdot I(X_1 \le \theta) \cdots \frac{1}{\theta} I(X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I(X_1 \le \theta) \cdots I(X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I(X_1 \le \theta, \dots, X_n \le \theta) \\ &= \left(\frac{1}{\theta}\right)^n \cdot I\left(\max_{i=1,\dots,n} X_i \le \theta\right) \end{aligned}$$

$\Rightarrow$ not differentiable with respect to θ





$$\Rightarrow \arg\max_{\theta} L(\theta) = \max_{i=1,\dots,n} X_i = \max\{X_1, \dots, X_n\}$$