Defn: Likelihood Function

If $x_n \overset{iid}{\sim} f_\theta$ where $\theta \in \Theta$.

Recall: joint distribution of data

$$\boxed{f_\theta(\underset{\sim}{x}) = \prod_{n=1}^{N} f_\theta(x_n)}$$

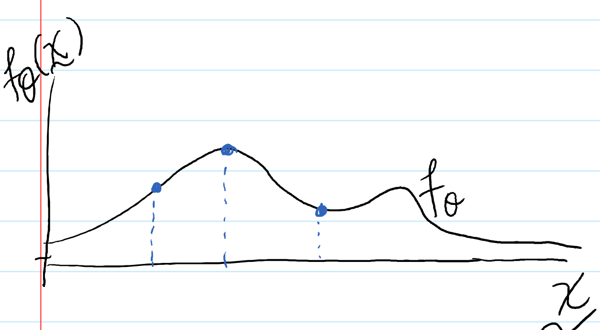
typically we think of this as a function of $\underset{\sim}{x}$

the likelihood function is viewing this joint dist. as a function of

Two ways of thinking about joint

Way 1

$f_\theta : \mathbb{R}^N \to \mathbb{R}$



Way 2: Likelihood function

$L : \Theta \to \mathbb{R}$

$L(\theta) = f_\theta(\underset{\sim}{x})$

Often its useful to consider the log-likelihood fn

$$\ell(\theta) = \log(L(\theta))$$

log-likelihood fn

natural logarithm

Defn: Maximum Likelihood Estimator (MLE)

Idea: want to estimate θ.

Choose $\hat{\theta}$ that gives the largest likelihood.

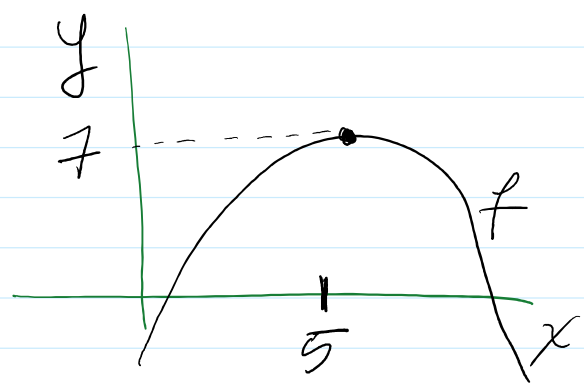
$$\hat{\theta}_{MLE}(\underset{\sim}{x}) = \hat{\theta}_{MLE} \overset{\text{def}}{=} \arg\max_{\theta \in \Theta} L(\theta)$$

argument that maximizes $L(\theta)$

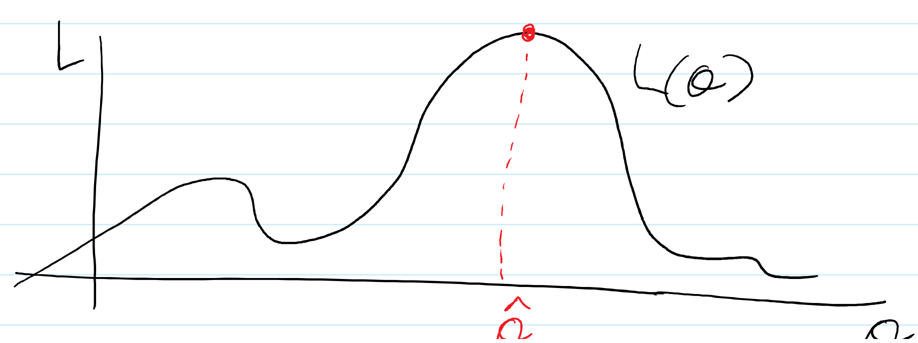
Aside: max v. argmax

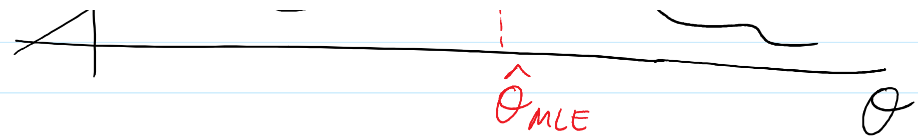
$$\max_x f(x) = 7$$

$$\arg\max_x f(x) = 5$$



MLE pictorally



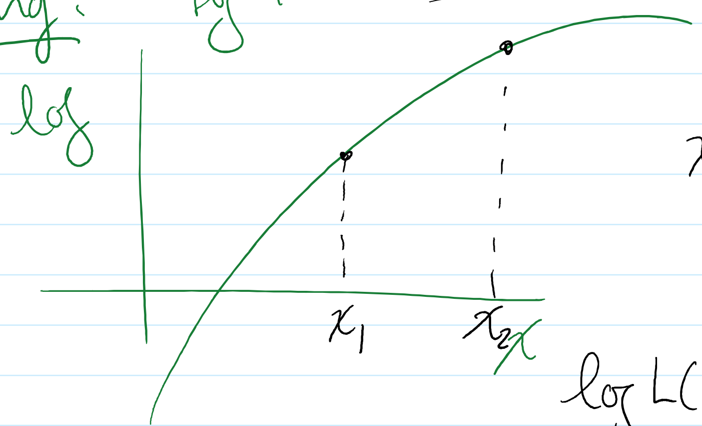


Why $\ell(\theta)$ – log likelihood function?

Fact:

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} \ell(\theta)$$

Why? log is an increasing function



$$x_2 > x_1 \Rightarrow \log(x_2) > \log(x_1)$$

$$L(\hat{\theta}_{MLE}) \geq L(\theta) \quad \forall \theta \in \Theta$$

$$\log L(\hat{\theta}) = \ell(\hat{\theta}) \geq \ell(\theta) = \log L(\theta)$$
$$\theta \in \Theta$$

Ex. $X_n \overset{iid}{\sim} N(\theta, 1)$ and $\theta \in \mathbb{R}$

What is the MLE of θ?

(1) find the likelihood function $L(\theta)$

$$L(\theta) = f_\theta(\underset{\sim}{x}) = \prod_{n=1}^{N} f_\theta(x_n) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_n-\theta)^2\right)$$

$$= (2\pi)^{-N/2} \exp\left(-\frac{1}{2}\sum_{n=1}^{N}(x_n-\theta)^2\right)$$

$$e^{a_1} e^{a_2} \cdots e^{a_n}$$
$$= e^{a_1+a_2+a_3+\cdots+a_n}$$

$$\ell(\theta) = \log L(\theta)$$

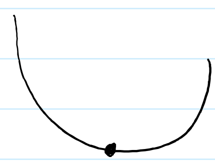
$$\ell(\theta) = \log L(\theta)$$
$$= \log\left((2\pi)^{-N/2}\right) - \frac{1}{2}\sum_{n=1}^{N}(x_n - \theta)^2$$
$$= -\frac{N}{2}\log(2\pi) - \frac{1}{2}\sum_{n=1}^{N}(x_n - \theta)^2$$

$$e^{a_1}e^{a_2}\cdots e^{a_n} = e^{a_1+a_2+a_3+\cdots+a_n}$$

Recall: $\hat{\theta}_{MLE} = \arg\max_{\theta} \ell(\theta)$

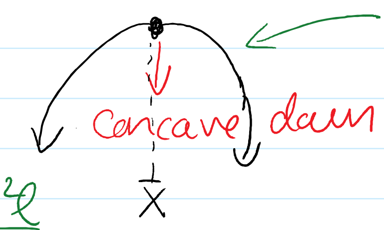
(2) take a derivative wrt θ and set $= 0$ and solve for θ

$$\frac{\partial \ell}{\partial \theta} = -\frac{1}{2}\sum_{n=1}^{N} 2(x_n - \theta)(-1) = \sum_{n=1}^{N}(x_n - \theta)$$
$$= \sum_{n=1}^{N} x_n - N\theta = 0$$



$$N\theta = \sum_{n=1}^{N} x_n$$

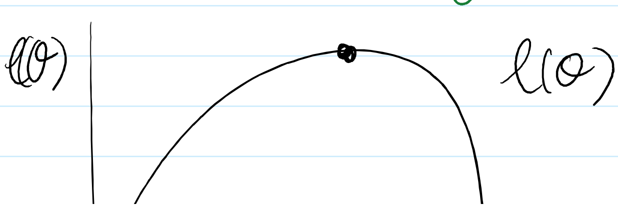
or $\boxed{\hat{\theta}_{MLE} = \frac{1}{N}\sum_{n=1}^{N} x_n = \overline{X}}$

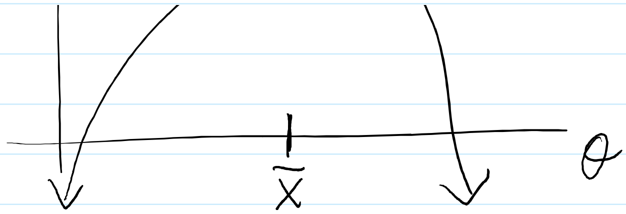


Technically we should check $\frac{\partial^2 \ell}{\partial \theta^2}$

$\rightarrow$ $\frac{\partial^2 \ell}{\partial \theta^2} = -N < 0$ so ℓ is concave down ✓

$\rightarrow$ $\lim_{\theta \to \pm\infty} L(\theta) = 0$ and $\lim_{\theta \to \pm\infty} \ell(\theta) = -\infty$





Theorem: MLEs are based off of sufficient stats.

$$\hat{\theta}_{MLE} = \text{function}(T)$$

T sufficient for θ.

pf Factorization theorem

$$L(\theta) = f_\theta(\underset{\sim}{x}) = h(\underset{\sim}{x})\, g(\theta, T)$$

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta)$$

$$= \arg\max_{\theta} h(\underset{\sim}{x})\, g(\theta, T)$$

$$= \arg\max_{\theta} g(\theta, T)$$

can ignore b/c it doesn't depend on θ

function of T

$$\boxed{\begin{aligned} &\arg\max_{x\in[1,2]} 5x^2 = 2 \\ &\max_{x\in[1,2]} 5x^2 = 5\cdot 4 \end{aligned}}$$

Ex. let $X_n \overset{iid}{\sim} \text{Bernoulli}(p)$

where $p \in [0,1]$

what is $\hat{p}_{MLE}$?

$$f(x) = \begin{cases} p & x=1 \\ 1-p & x=0 \end{cases}$$

$$= p^x (1-p)^{1-x}$$

what is $\hat{p}_{MLE}$?

$= p^{\cdot}(1-p)^{\cdot}$

① Write $L(p)$ or $\ell(p)$

$$L(p) = f_p(\underset{\sim}{x}) = \prod_{n=1}^{N} p^{x_n}(1-p)^{1-x_n}$$

$$= p^{\sum_n x_n}(1-p)^{\sum(1-x_n)}$$

$$= p^{\sum x_n}(1-p)^{N-\sum_n x_n}$$

$$\boxed{\prod_i a^{b_i} = a^{\sum b_i}}$$

$$\ell(p) = \log L(p) = \left(\sum_n x_n\right)\log(p) + \left(N - \sum_n x_n\right)\log(1-p)$$

$$\boxed{\begin{aligned} \log(ab) &= \log(a) + \log(b) \\ \log(a^b) &= b\log(a) \end{aligned}}$$

② $\dfrac{\partial \ell}{\partial p} = 0$

$$\frac{\partial \ell}{\partial p} = \frac{\sum_n x_n}{p} - \frac{N - \sum_n x_n}{1-p} = 0$$

$$\Rightarrow (1-p)\sum_n x_n - p\left(N - \sum_n x_n\right) = 0$$

$$\Rightarrow \sum_n x_n - \cancel{p\sum x_n} - pN + \cancel{p\sum_n x_n} = 0$$

$$\Rightarrow \sum_n x_n = pN$$

$$\Rightarrow \boxed{\hat{p}_{MLE} = \frac{1}{N}\sum_n x_n = \overline{x}}$$

sum is # of 1s

$$\frac{\text{sum is \# of 1s}}{N} = \% \text{ of 1s}$$

odds: $\boxed{\eta = \frac{p}{1-p}}$

Notice: $(1-p)\eta = p$

$$\Rightarrow \eta - \eta p = p \Rightarrow \eta = \eta p + p = \eta = (1+\eta)p$$

$$\Rightarrow \boxed{p = \frac{\eta}{1+\eta}}$$

What is the MLE for η? $\hat{\eta}_{MLE}$?

Could find this by re-parameterizing in terms of η.

$$L(p) = p^{\sum x_n}(1-p)^{N-\sum x_n}$$

1 $$L(\eta) = \left(\frac{\eta}{1+\eta}\right)^{\sum x_n}\left(1-\frac{\eta}{1+\eta}\right)^{N-\sum x_n}$$

$$1-\frac{\eta}{1+\eta} = \frac{1+\eta-\eta}{1+\eta} = \frac{1}{1+\eta}$$

$$\ell(\eta) = (\sum x_n)\log\left(\eta/_{1+\eta}\right) + (N-\sum x_n)\log\left(1/_{1+\eta}\right)$$

$$\log(a/b) = \log(a) - \log(b)$$

$$= (\sum x_n)\left[\log(\eta) - \log(1+\eta)\right] - (N-\sum x_n)\log(1+\eta)$$

$$= (\sum x_n)\log(\eta) - \cancel{(\sum x_n)\log(1+\eta)} - N\log(1+\eta) + \cancel{(\sum x_n)\log(1+\eta)}$$

$$= \left(\sum x_n\right)\log(\eta) - \cancel{\left(\sum x_n\right)\log(1+\eta)} - N\log(1+\eta) + \cancel{\left(\sum x_n\right)\log(1+\eta)}$$

$$= \left(\sum x_n\right)\log(\eta) - N\log(1+\eta)$$

2 $\boxed{\frac{\partial \ell}{\partial \eta} = 0}$

$$\frac{\partial \ell}{\partial \eta} = \frac{\sum x_n}{\eta} - \frac{N}{1+\eta} = 0$$

$$\Rightarrow (1+\eta)\left(\sum x_n\right) - \eta N = 0$$

$$\Rightarrow \left(\sum x_n\right) = -\eta\left(\sum x_n\right) + \eta N = \eta\left(N - \sum x_n\right)$$

$$\Rightarrow \boxed{\hat{\eta}_{MLE} = \frac{\frac{1}{N}\sum x_n}{\frac{1}{N}N - \frac{1}{N}\sum x_n} = \frac{\bar{X}}{1-\bar{X}}}$$

Recall that $\boxed{\eta = \frac{p}{1-p}}$ and $\boxed{\hat{p}_{MLE} = \bar{X}}$

notice: $\hat{\eta}_{MLE} = \frac{\hat{p}}{1-\hat{p}}$.

Punchline: we can do this generally.

Theorem: Transformation for MLEs.

If $\hat{\theta}$ is the MLE for θ then for any function τ, the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$.

Ex. We showed that

If $X_n \overset{iid}{\sim} N(\theta, 1)$ then $\hat{\theta}_{MLE} = \bar{X}$.

What is the MLE for θ^2?

It is $\bar{X}^2$.

Interpretation of MLE

$$L(\theta) = L(\theta | \underset{\sim}{X}) = f_\theta(\underset{\sim}{x})$$

given that I observed some data $\underset{\sim}{x}$

$L(\theta)$ says how "likely" it is that this observed data came from a dist w/ parameter θ.

$$L(1 | \underset{\sim}{X}) \quad \text{v.s.} \quad L(2 | \underset{\sim}{X})$$

likelihood of observing $\underset{\sim}{X}$ if my parameter $\theta = 1$

" " " if my parameter is $\theta = 2$

MLEs: estimate θ as the value of my parameter that is most likely

parameter that is most likely
given my data.

$X_n \sim N(\theta, 1)$

$\underset{\sim}{X} = (1, 3, 7, 2, 4)$

$\theta = 10{,}000$ unlikely
$\theta = 4$ more likely