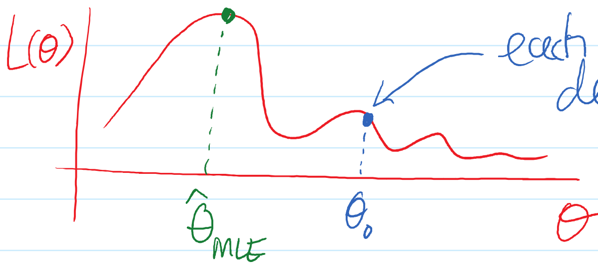
Lecture 6 - More MLEs

Recap: $L(\theta) = f_\theta(\underset{\sim}{x})$; $\ell(\theta) = \log L(\theta)$



each pt describes for a fixed $\underset{\sim}{x}$ the "likelihood" of observing $\underset{\sim}{x}$ given that $\theta = \theta_0$

Ex. $X_n \overset{iid}{\sim} Pois(\lambda)$

① $$L(\lambda) = f_\lambda(\underset{\sim}{x}) = \prod_{n=1}^{N} \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} = \frac{e^{-N\lambda}\lambda^{\sum_{n=1}^{N} x_n}}{\prod_{n=1}^{N} x_n!}$$

$$\ell(\lambda) = \log L(\lambda) = -N\lambda + \left(\sum_{n=1}^{N} x_n\right)\log(\lambda) - \sum_{n=1}^{N}\log(x_n!)$$

② $$\frac{\partial \ell}{\partial \lambda} = -N + \frac{\sum_{n=1}^{N} x_n}{\lambda} = 0$$

then $\boxed{\hat{\lambda} = \overline{X}}$

Ex. $X_n \overset{iid}{\sim} Exp(\lambda)$ recall $E[X_n] = 1/\lambda$

$$f(x) = \lambda e^{-\lambda x} \mathbb{1}(x>0)$$

1. $$L(\lambda) = \prod_{n=1}^{N} f(\lambda) = \prod_{n=1}^{N} \lambda e^{-\lambda x_n} \mathbb{1}(x_n > 0)$$

$$= \lambda^N e^{-\lambda \sum_{n=1}^{N} x_n} \prod_{n=1}^{N} \mathbb{1}(x_n > 0)$$

$$= \lambda^N e^{-\lambda \sum_{n=1}^{N} x_n} \mathbb{1}(X_{(1)} > 0)$$

$\mathbb{1}(A)\mathbb{1}(B) = \mathbb{1}(A \text{ and } B)$

$$\ell(\lambda) = N \log \lambda - \lambda \sum_{n=1}^{N} x_n + \log \mathbb{1}(X_{(1)} > 0)$$

rewrite $\mathbb{1}(X_{(1)} > 0)$

2. $$\frac{\partial \ell}{\partial \lambda} = \frac{N}{\lambda} - \sum_{n=1}^{N} x_n = 0$$

$$\boxed{\hat{\lambda}_{MLE} = \frac{1}{\bar{X}}}$$

$E[\bar{X}] = \frac{1}{\lambda}$

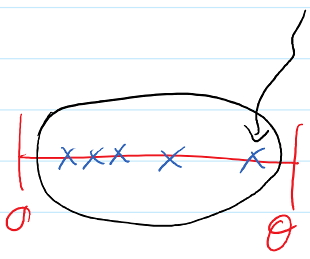
If $X_n \overset{iid}{\sim} f_\beta$ when $f_\beta(x) = \frac{1}{\beta} e^{-x/\beta} \mathbb{1}(x>0)$

$E[X_n] = \beta$

by transformation prop of MLES $\boxed{\lambda = 1/\beta}$

Since $\beta = 1/\lambda$ then $\hat{\beta}_{MLE} = 1/\hat{\lambda}_{MLE} = \bar{X}$

Ex. $X_n \overset{iid}{\sim} \text{Unif}(0,\theta)$ where $\theta > 0$



$$f_\theta(x) = \frac{1}{\theta}\,\mathbb{1}(0 < x < \theta)$$

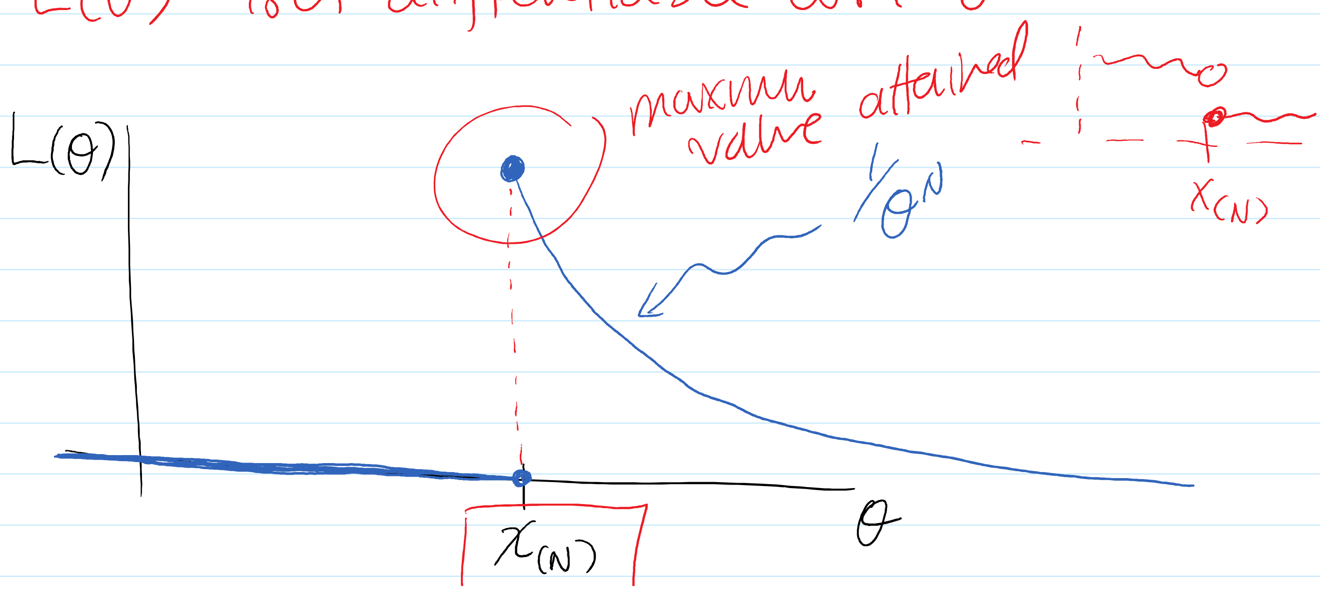
support depends on θ

(1) $$L(\theta) = \prod_{n=1}^{N} f_\theta(x_n) = \prod_{n=1}^{N} \frac{1}{\theta}\underbrace{\mathbb{1}(0 < x_n < \theta)}_{\mathbb{1}(x_n > 0)\,\mathbb{1}(x_n < \theta)}$$

$$= \theta^{-N}\underbrace{\prod_{n=1}^{N}\mathbb{1}(x_n > 0)}_{\mathbb{1}(X_{(1)} > 0)}\;\underbrace{\prod_{n=1}^{N}\mathbb{1}(x_n < \theta)}_{\mathbb{1}(X_{(N)} < \theta)}$$

$$= \theta^{-N}\,\mathbb{1}(X_{(1)} > 0)\,\underbrace{\mathbb{1}(X_{(N)} < \theta)}$$

$L(\theta)$ isn't differentiable wrt θ



$$\boxed{\hat{\theta} = X_{(N)}}$$

Ex. $X_n \overset{iid}{\sim} N(\theta, 1)$ where $\theta \geq 0$

To find MLE it must satisfy the constraint

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

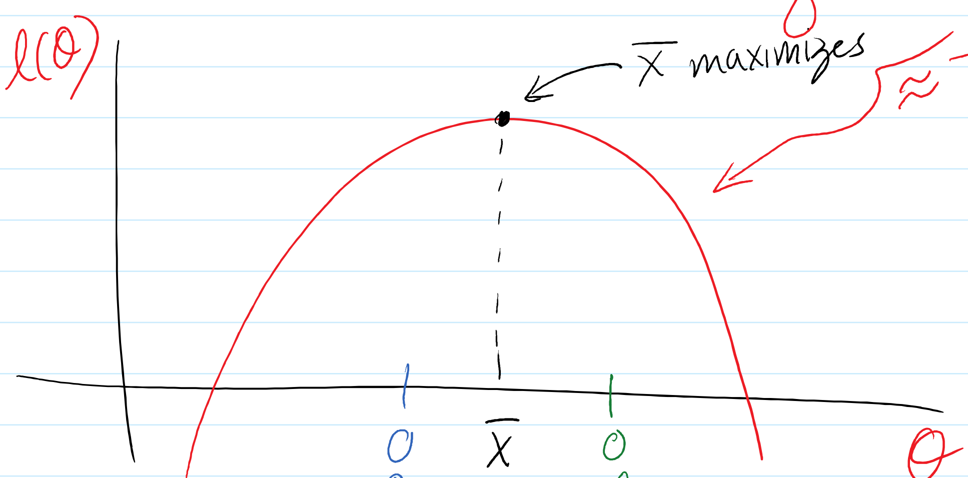
$\hat{\theta} = \underset{\theta \in \Theta}{\operatorname{argmin}}\, L(\theta)$

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

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

a quadratic in θ

$\approx -\theta^2$

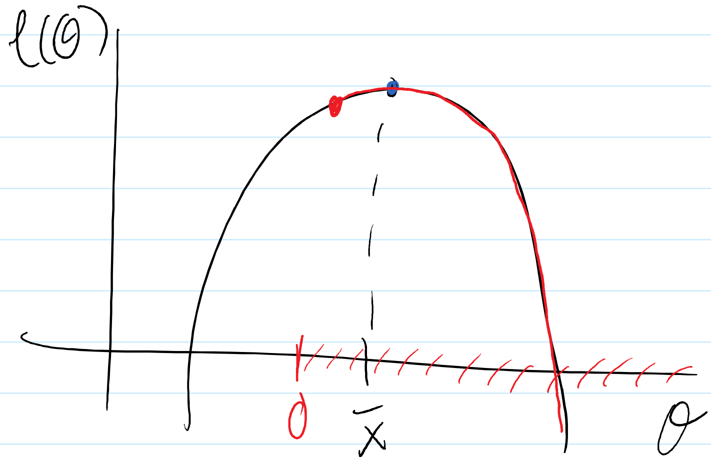


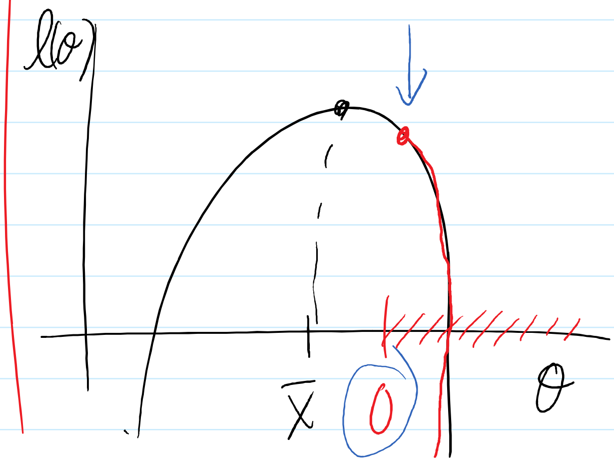
two situations: (1) $\bar{X} > 0$ or (2) $\bar{X} < 0$

$$\boxed{\hat{\theta} = \bar{X}}$$

$$\boxed{\hat{\theta} = 0}$$

$\ell(\theta)$





All together

$$\hat{\theta} = \begin{cases} 0, & \bar{x} < 0 \\ \bar{x}, & \bar{x} > 0 \end{cases}$$

Ex,

$X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both unknown $\mu \in \mathbb{R}$ and $\sigma^2 > 0$

1

$$L(\mu, \sigma^2) = \prod_{n=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_n - \mu)^2\right)$$

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

$$\ell(\mu, \sigma^2) = -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\underbrace{\sigma^2}_{\tau}) - \frac{1}{2\underbrace{\sigma^2}_{\tau}}\sum_{n=1}^{N}(x_n-\mu)^2$$

(2) $\boxed{\frac{\partial \ell}{\partial \mu} = 0 \quad \text{and} \quad \frac{\partial \ell}{\partial \sigma^2} = 0}$

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

$$= \frac{1}{\sigma^2}\left[\left(\sum_{n=1}^{N} x_n\right) - N\mu\right] = 0$$

$$\Rightarrow \sum_{n=1}^{N} x_n = N\mu \Rightarrow \boxed{\hat{\mu}_{MLE} = \bar{x}}$$

$\tau = \sigma^2$

$$\frac{\partial \ell}{\partial \sigma^2} = \frac{\partial \ell}{\partial \tau} = -\frac{N}{2\tau} + \frac{1}{2\tau^2}\sum_{n=1}^{N}(x_n-\hat{\mu})^2 = 0$$

$$\Rightarrow -1 + \frac{1}{\tau}\frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})^2 = 0$$

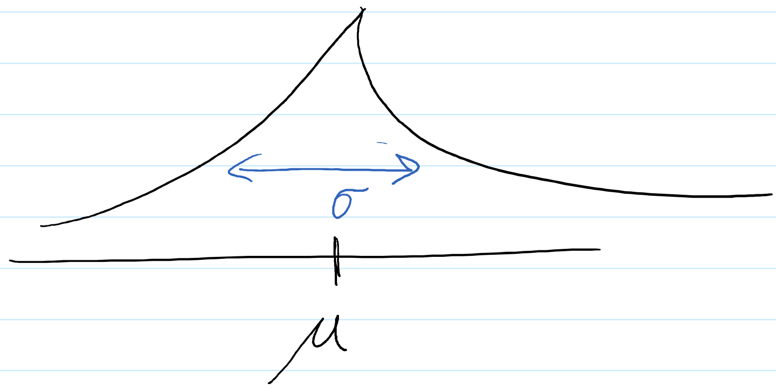
$$\Rightarrow \boxed{\hat{\tau} = \hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{n=1}^{N}(x_n-\bar{x})^2}$$

looks familar!

almost $s^2 = \frac{1}{N-1}\sum_{n=1}^{N}(x_n-\bar{x})^2$

Ex. $X_n \overset{iid}{\sim} \text{Laplace}(\mu, \sigma)$

$$f(x) = \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|x-\mu|\right)$$



MLE?

(1) $$L(\mu,\sigma) = \prod_{n=1}^{N} \frac{1}{2\sigma} \exp\left(-\frac{1}{\sigma}|x_n-\mu|\right)$$

$$= 2^{-N}\sigma^{-N} \exp\left(-\frac{1}{\sigma}\sum_{n=1}^{N}|x_n-\mu|\right)$$

(2) $$\frac{\partial \ell}{\partial \mu} = 0 \quad \text{and} \quad \frac{\partial \ell}{\partial \sigma} = 0 \quad \checkmark$$

not diffentable

$$\ell(\mu,\sigma) = -N\log 2 - N\log\sigma - \frac{1}{\sigma}\sum_{n=1}^{N}|x_n-\mu|$$

$$\frac{\partial \ell}{\partial \sigma} = -\frac{N}{\sigma} + \frac{1}{\sigma^2}\sum_{n=1}^{N}|x_n-\mu| = 0$$

then $\boxed{\hat{\sigma} = \frac{1}{N}\sum^{N}|x_n-\hat{\mu}|}$

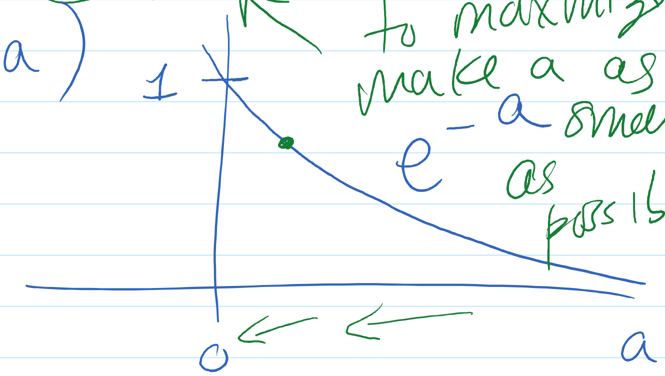
then $\boxed{\hat{\sigma} = \frac{1}{N}\sum_{n=1}^{N} |x_n - \hat{\mu}|}$

$\hat{\mu}$?

$$L(\mu,\sigma) \propto \exp\Big(-\frac{1}{\sigma}\underbrace{\overbrace{\sum_{n=1}^{N} |x_n-\mu|}^{>0}}\Big)$$

$\exp(-a)$

$a>0$



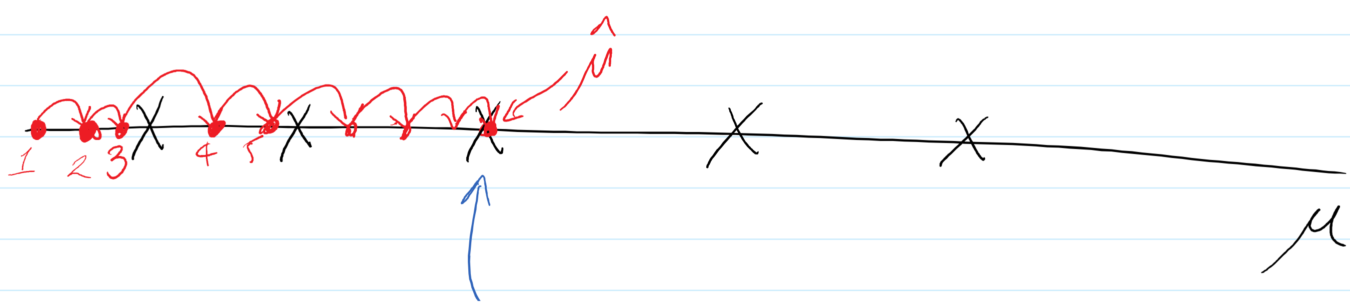
to maximize make a as small as possible

Make $\sum_{n=1}^{N} |x_n-\mu|$ as small as possible.

Fact: ① $\sum_{n=1}^{N} (x_n-\mu)^2$ minimized if $\mu = \bar{x}$
(pf. w/ calc I)

② $\underbrace{\sum_{n=1}^{N} |x_n-\mu|}$ minimized if $\boxed{\hat{\mu} = \text{median}(\underset{\sim}{x})}$

why? total dist of x_ns to μ



median

μ