Ex. Let $X_n \overset{iid}{\sim} \text{Exp}(\lambda)$, $\lambda > 0$.

Let's get MLE for λ.

① Get $L(\lambda)$ and $\ell(\lambda)$

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

$$= \lambda^N e^{-\lambda \sum_n x_n} \mathbb{1}(x_{(1)} > 0)$$

$$\ell(\lambda) = \log L(\lambda) = N \log(\lambda) - \lambda \sum_n x_n + \log \mathbb{1}(x_{(1)} > 0)$$

② Set $\frac{\partial \ell}{\partial \lambda} = 0$ solve for λ

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

So $\boxed{\hat{\lambda} = \frac{N}{\sum_n x_n} = \frac{1}{\bar{x}}}$

$\sum_n x_n$

Notice: $E[X_n] = 1/\lambda = E[\bar{X}]$

Note: $E[1/\bar{X}] \neq \lambda$

Sometimes parameterize exp. w/ the mean

$$\beta = 1/\lambda = E[X_n]$$

i.e. $f(x_n) = \frac{1}{\beta} e^{-x_n/\beta} \mathbb{I}(x_n > 0)$

What's the MLE for β?

By transf. thm, $\hat{\beta} = 1/\hat{\lambda} = \bar{X}$.

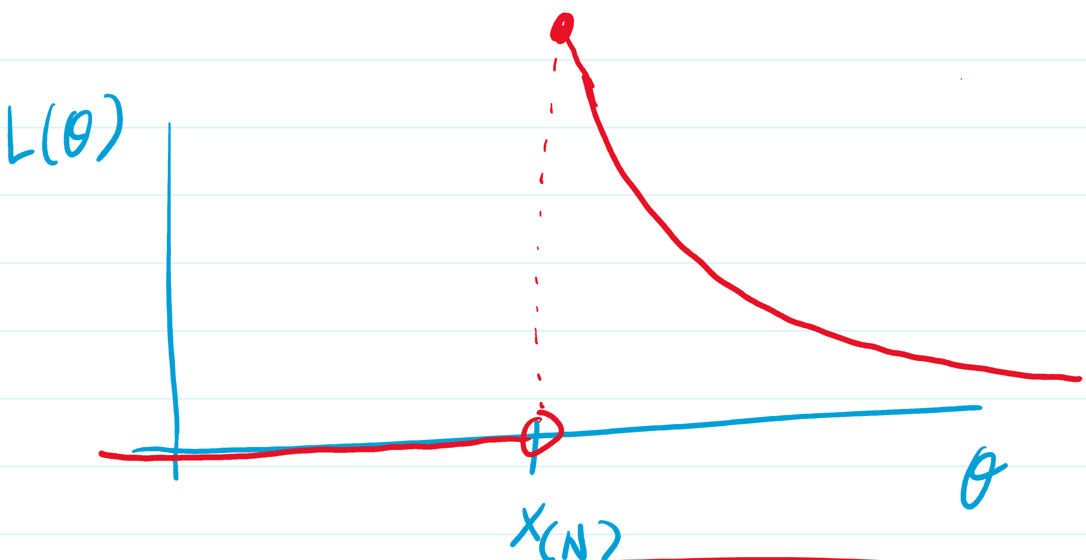
Ex. $X_n \overset{iid}{\sim} U(0, \theta)$ for $\theta > 0$.

What is the MLE?

$$L(\theta) = \prod_n f(x_n) = \prod_n \frac{1}{\theta} \mathbb{1}(0 \le x_n \le \theta)$$

$$= \theta^{-N} \prod_n \mathbb{1}(x_n \ge 0) \prod_n \mathbb{1}(x_n \le \theta)$$

$$= \theta^{-N} \mathbb{1}(X_{(1)} \ge 0) \mathbb{1}(X_{(N)} \le \theta)$$



$$\boxed{\hat{\theta} = \arg\max_{\theta} L(\theta) = X_{(N)}}.$$

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

Get $L(\theta)$:

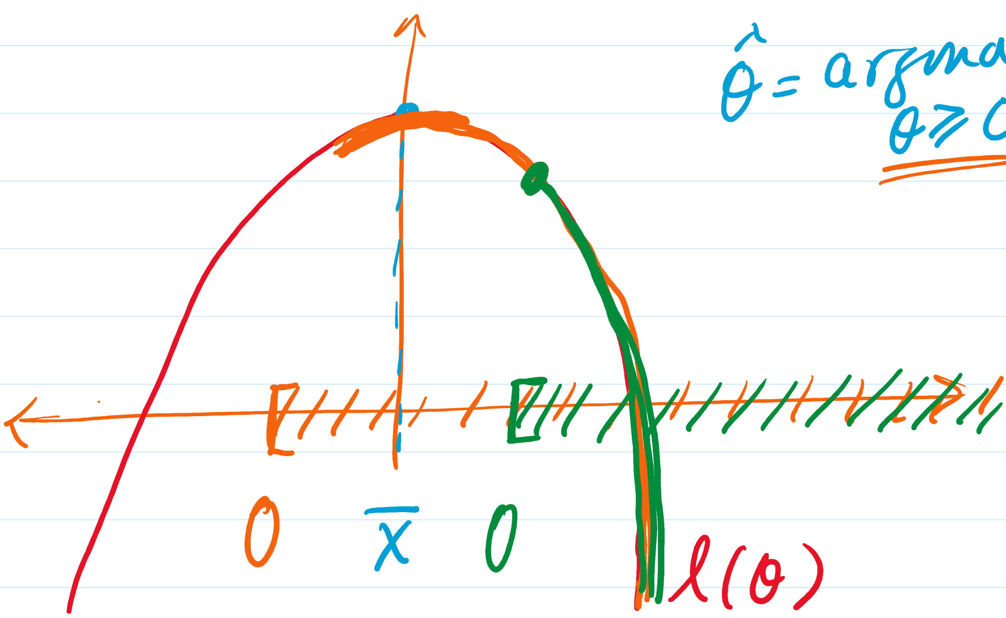
$$L(\theta) = \prod_n \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}(x_n - \theta)^2\right)$$

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

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

like $-\theta^2$

$$\hat{\theta} = \underset{\theta \geq 0}{\operatorname{argmax}}\, \ell(\theta)$$



Case 1: $\bar{x} \geq 0 \Rightarrow \hat{\theta} = \bar{x}$

Case 2: $\bar{x} < 0 \Rightarrow \hat{\theta} = 0$

Case 2: $\bar{X} < 0 \Rightarrow \hat{\theta} = 0$

$$\boxed{\hat{\theta} = \max(\bar{X}, 0)}$$

Ex. Let $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ both unknown.

Lets get MLE for μ and σ^2.

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

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

$$\ell(\mu, \sigma^2) = \log L(\mu, \sigma^2) \qquad \tau = \sigma^2$$

$$= -\frac{N}{2}\log(2\pi) - \frac{N}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_n (x_n - \mu)^2$$

$$\ell(\mu, \tau) = -N/2 \log(2\pi) - N/2 \log(\tau) - \frac{1}{2\tau} \sum_n (x_n - \mu)^2$$

② Set gradient to zero.

② Set gradient to zero.

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

$$= \frac{1}{\sigma^2}\sum_n (x_n-\mu) = 0$$

$$\Rightarrow \sum_n x_n = \sum_n \mu$$

$$\Rightarrow \sum_n x_n = N\mu$$

$$\Rightarrow \boxed{\hat{\mu} = \overline{X}}$$

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

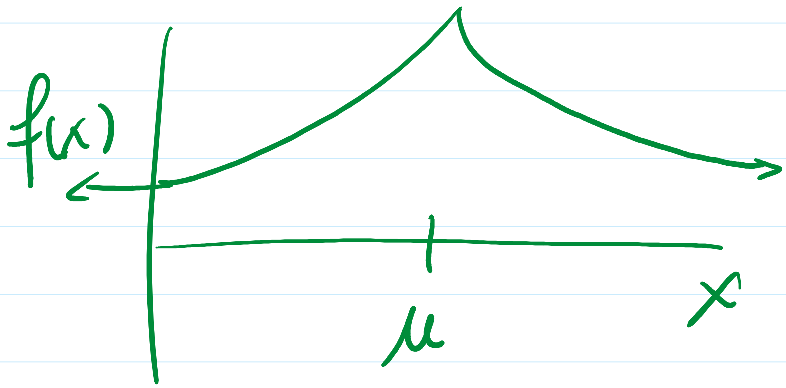
$$\Rightarrow -\frac{N}{\tau} + \frac{1}{\tau^2}\sum_n (x_n-\overline{X})^2 = 0$$

$$= \sum_n (x_n-\overline{X})^2 = N\tau$$

$$\Rightarrow \tau = \boxed{\frac{1}{N}\sum_n (x_n-\overline{X})^2 = \hat{\sigma}^2}$$

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

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



What's the MLEs ?

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

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

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

Problem: not diffable WRT μ.

Can look at $\frac{\partial \ell}{\partial \sigma} = \frac{-N}{\sigma} + \frac{1}{\sigma^2} \sum_n |X_n - \mu| = 0$

Solving for σ we get

$$\boxed{\hat{\sigma} = \frac{1}{N} \sum_n |X_n - \hat{\mu}|}$$

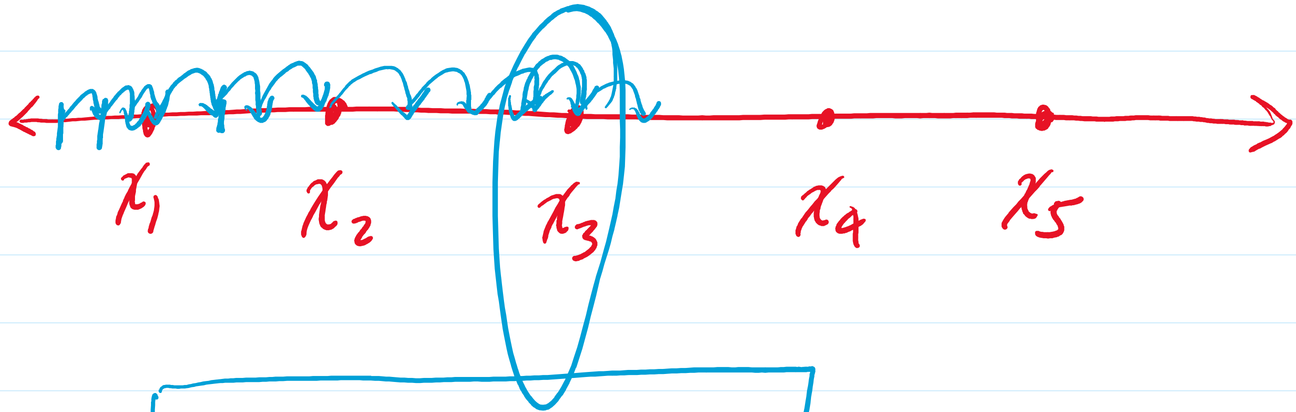
How to get $\hat{\mu}$?

fix σ

$$\ell(\mu, \sigma) \propto -\sum_n |X_n - \mu|$$

To max wrt μ we should minimize

$\sum_n |X_n - \mu| =$ total dist of X_n's to μ



$$\boxed{\hat{\mu} = \text{median}\{X_n\}}$$

Evaluation

Defn: Mean Squared Error (MSE)

If $X_n \overset{iid}{\sim} f_\theta$ where $\theta \in \Theta$ and let $\hat{\theta}$ be an est. of θ. The MSE of $\hat{\theta}$ est. θ is

$$MSE_\theta(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right]$$

If $\hat{\theta}$ is good at est. θ then MSE small.

Might prefer $\hat{\theta}$ w/ small MSE.

Defn: Bias The bias of $\hat{\theta}$ est. θ is

$$B_\theta(\hat{\theta}) = E[\hat{\theta} - \theta] = E[\hat{\theta}] - \theta$$

If $B(\hat{\theta}) > 0$ I tend to over-estimate
< 0 " under "

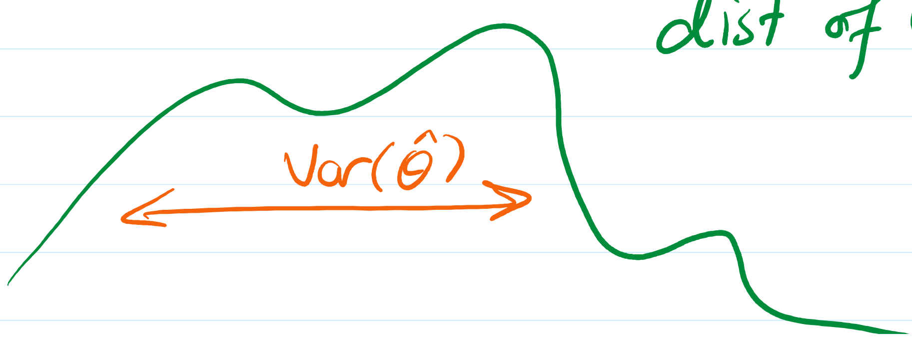
$B(\hat{\theta}) = 0$ we say $\hat{\theta}$ is unbiased.

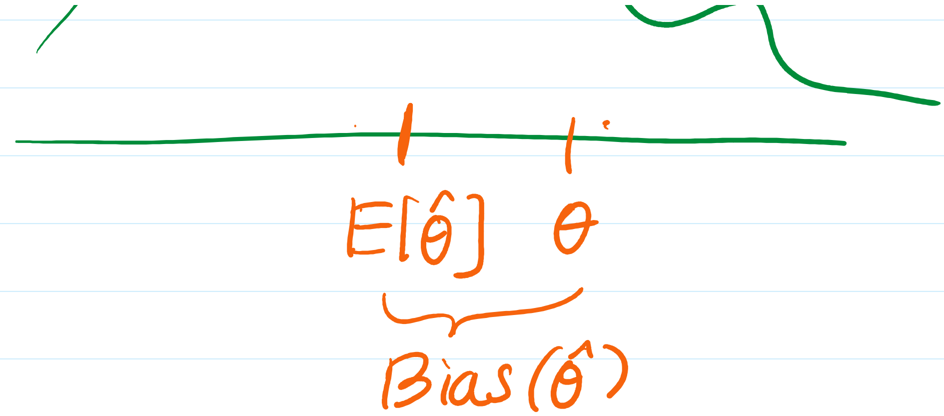
Variance:

Recall $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$, so $\hat{\theta}$ is random,
so it has a variance: $Var(\hat{\theta})$.

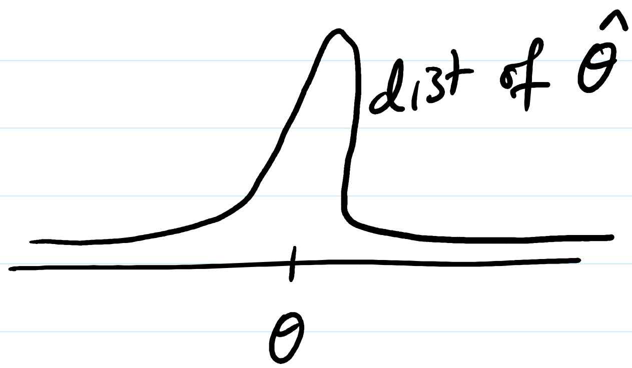
Ex.

dist of $\hat{\theta}$



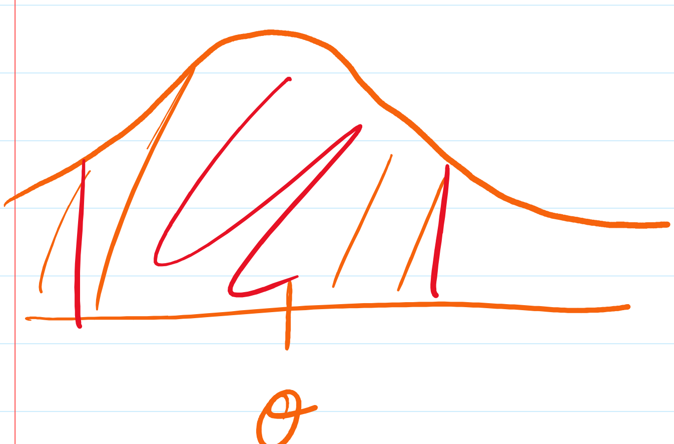


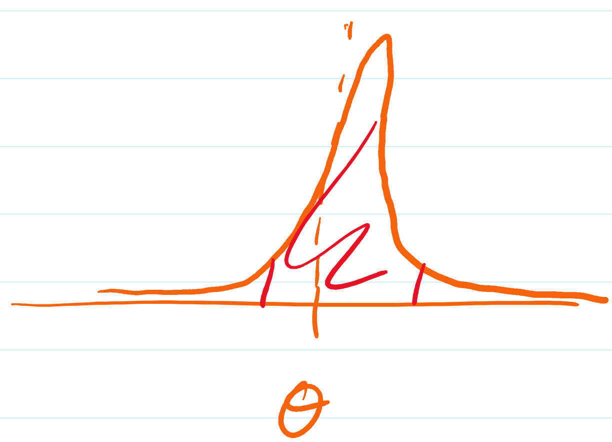
Ideally: small bias and variance



Sometimes, non-zero bias is better:

(1) high var, no bias (2) low var, some bias





Theorem: MSE = bias^2 + Var.

pf. $MSE(\hat{\theta}) = E[(\hat{\theta}-\theta)^2]$

$$= E\Big[\big(\underbrace{\hat{\theta} - E[\hat{\theta}]}_{a} + \underbrace{E[\hat{\theta}] - \theta}_{b}\big)^2\Big]$$

$(a+b)^2 = a^2 + b^2 + 2ab$

$$= E[a^2] + E[b^2] + 2E[ab]$$

...