what value of μ makes

$$J = \sum_n (x_n - \mu)^2$$ as small as possible?

$\mu = \bar{x}$; $\frac{dJ}{\partial \mu} = 0 \ldots$

Soln: take derivative wrt μ.

Our problem: $T = \sum_n |x_n - \mu|$ as small as possible

$$\boxed{\mu = \text{median}(\underset{\sim}{x})}$$

We've talked about

(1) MoM

(2) MLE

How do we compare estimators?

Defn: Mean Squared Error (MSE)

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

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

Let $\hat{\theta}$ be an estimater of θ.

We define the MSE of $\hat{\theta}$ estimating θ as

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

avg. sq. dist. of $\hat{\theta}$ from θ

If our estimater $\hat{\theta}$ is "good" the MSE is small.

<u>Idea:</u> If I have $\hat{\theta}_1$ and $\hat{\theta}_2$ I could say the better estimater is the one w/ a smaller MSE.

<u>Defn: Bias</u>

The bias of $\hat{\theta}$ in estimating θ is

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

on avg. does $\hat{\theta}$ over/under-estimate θ

We say an estimater is unbiased if

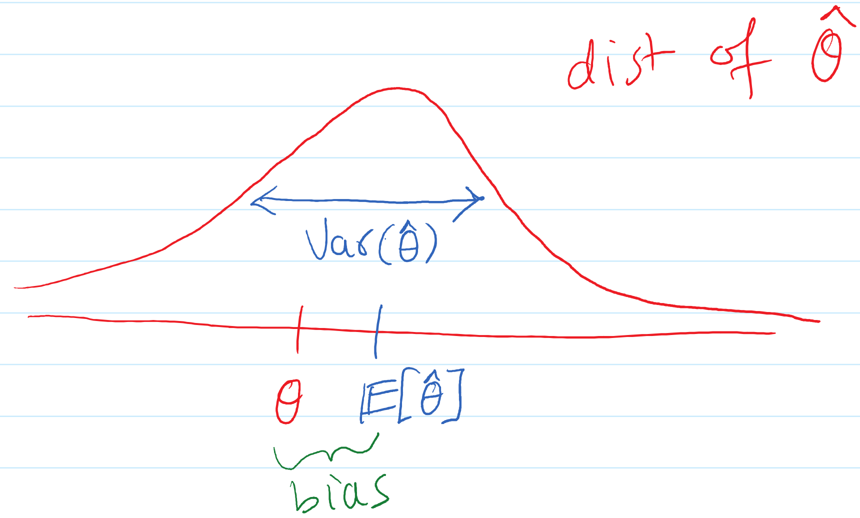
We say an estimater is unbiased if

$$B_\theta(\hat{\theta}) = 0 \quad \text{i.e. } \mathbb{E}[\hat{\theta}] = \theta$$

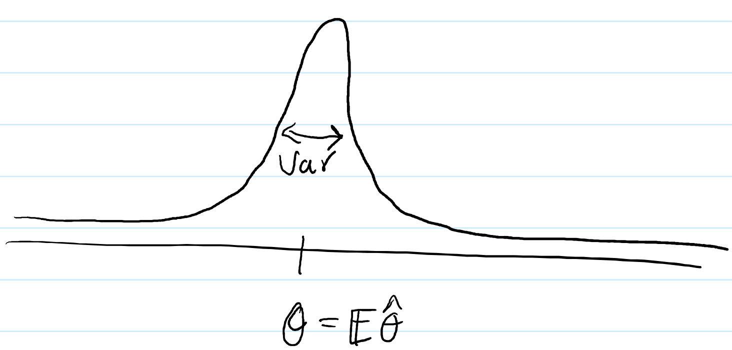
Defn: Variance

Recall $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$ and so it has a variance $\mathrm{Var}_\theta(\hat{\theta})$.

Ex.



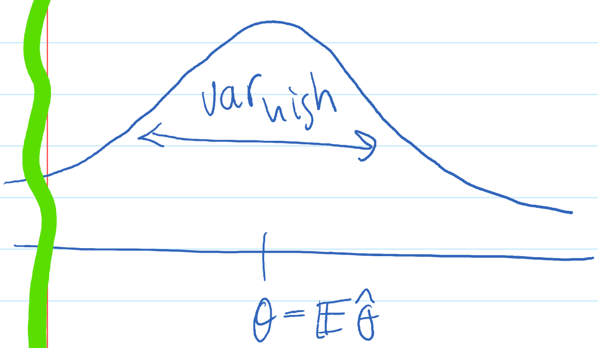
Ideally: $B_\theta(\hat{\theta})$ is zero (or small) and so is the variance



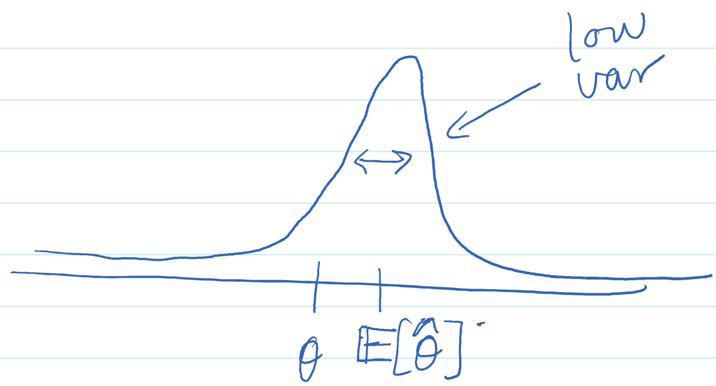
$\theta = \mathbb{E}\hat{\theta}$

Sometimes Unbiased est. aren't the best:

Unbiased $\hat{\theta}$



biased $\hat{\theta}$



Theorem: MSE = bias^2 + Var

$$MSE(\hat{\theta}) = B(\hat{\theta})^2 + Var(\hat{\theta}).$$

sq. scale

on linear scale

Pf.

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

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

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

$$= \mathbb{E}[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2 + (\mathbb{E}\hat{\theta}-\theta)^2 + 2(\hat{\theta}-\mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta}-\theta)]$$

not random

$$= \mathbb{E}[(\hat{\theta}-\mathbb{E}\hat{\theta})^2] + \cancel{\mathbb{E}[}(\mathbb{E}\hat{\theta}-\theta)^2\cancel{]} + 2(\mathbb{E}\hat{\theta}-\theta)\mathbb{E}[\hat{\theta}-\mathbb{E}\hat{\theta}]$$

$$= \underbrace{\mathbb{E}[(\hat\theta - \mathbb{E}\hat\theta)^2]}_{Var(\hat\theta)} + \underbrace{\cancel{\mathbb{E}}[(\mathbb{E}\hat\theta - \theta)^2]}_{B(\hat\theta)^2} + 2(\mathbb{E}\hat\theta - \theta)\,\mathbb{E}[\hat\theta - \mathbb{E}\hat\theta]$$

$\mathbb{E}[\hat\theta] - \mathbb{E}[\hat\theta] = 0$

$Var(X) = \mathbb{E}[(X - \mathbb{E}X)^2]$	$B(\hat\theta) = \mathbb{E}\hat\theta - \theta$

<u>Ex.</u> $X_n \overset{iid}{\sim} f$

$$\mu = \mathbb{E}X_n \; ; \; \sigma^2 = Var(X_n)$$

Consider $\hat\mu = \bar X$

Showed prev: ① $\mathbb{E}\hat\mu = \mathbb{E}\bar X = \mu$ ← unbiased for μ

② $Var\,\hat\mu = Var\,\bar X = \sigma^2/N$

$$MSE(\hat\mu) = B(\hat\mu)^2 + Var(\hat\mu)$$
$$= 0 + \sigma^2/N$$
$$= \sigma^2/N$$

$\mathbb{E}\hat\mu - \mu$

If est. is unbiased $MSE = Var$.

Consider $S^2_{N-1} = \frac{1}{N-1}\sum_{n=1}^{N}(X_n - \bar X)^2$

We showed: $\mathbb{E}[S^2] = \sigma^2$

Facts:

If $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ then

$$\frac{N-1}{\sigma^2} S^2 \sim \chi^2(N-1)$$

Facts:

$$\begin{cases} Z \sim \chi^2(k) \\ \mathbb{E} Z = k \\ \operatorname{Var} Z = 2k \end{cases}$$

$$\operatorname{Var}\left(\frac{N-1}{\sigma^2} S^2\right) = 2(N-1)$$

so $$\frac{(N-1)^2}{\sigma^4} \operatorname{Var}(S^2) = 2(N-1)$$

hence $$\operatorname{Var}(S^2) = \frac{2\sigma^4}{N-1}$$

So $$\begin{aligned} \operatorname{MSE}(S^2) &= B(S^2)^2 + \operatorname{Var}(S^2) \\ &= 0 + \frac{2\sigma^4}{N-1} \end{aligned}$$

Consider $$\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^{N} (X_n - \bar{X})^2 = \frac{N-1}{N} S^2$$

$$S^2 = \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \bar{X})^2$$

Bias:

$$\begin{aligned} \operatorname{Bias}(\hat{\sigma}^2) = B\left(\frac{N-1}{N} S^2\right) &= \mathbb{E}\left[\frac{N-1}{N} S^2\right] - \sigma^2 \\ &= \frac{N-1}{N} \mathbb{E}[S^2] - \sigma^2 \end{aligned}$$

$$= \frac{N-1}{N}\mathbb{E}[S^2] - \sigma^2$$

$$= \frac{N-1}{N}\sigma^2 - \sigma^2$$

$$= -\frac{\sigma^2}{N}$$

Variance:

$$Var(\hat{\sigma}^2) = Var\left(\frac{N-1}{N}S^2\right)$$

$$= \frac{(N-1)^2}{N^2} Var(S^2)$$

$$= \frac{(N-1)^2}{N^2}\frac{2\sigma^4}{N-1} = \frac{2(N-1)}{N^2}\sigma^4$$

$$MSE(\hat{\sigma}^2) = Bias^2 + Var$$

$$= \left(\frac{-\sigma^2}{N}\right)^2 + \frac{2(N-1)}{N^2}\sigma^4$$

$$= \frac{\sigma^4}{N^2} + \frac{2(N-1)}{N^2}\sigma^4 = \boxed{\frac{2N-1}{N^2}\sigma^4}$$

$$MSE(S^2) = \boxed{\frac{2\sigma^4}{N-1}}$$

$$MSE(\hat{\sigma}^2) = \frac{2N-1}{N^2}\sigma^4 = \frac{2N-1}{N^2}\left(\frac{N-1}{2}\right)\left(\frac{2}{N-1}\right)\sigma^4$$

$$MSE(\hat{\sigma}^2) = \frac{2N-1}{N^2}\sigma^4 = \underbrace{\frac{2N-1}{N^2}\left(\frac{N-1}{2}\right)}\underbrace{\left(\frac{2}{N-1}\right)\sigma^4}_{MSE(S^2)}$$

$$\frac{(2N-1)(N-1)}{2N^2} = \frac{2N^2-3N+1}{2N^2} < 1$$

So $MSE(\hat{\sigma}^2) < MSE(S^2)$.

More generally: is there some $c \in \mathbb{R}$, that minimizes MSE cS^2.

$$\begin{cases} c = 1 \Rightarrow S^2 \\ c = \frac{N-1}{N} \Rightarrow \hat{\sigma}^2 \end{cases}$$

$$MSE(cS^2)$$

$$= Bias(cS^2)^2 + Var(cS^2)$$

$$= (\mathbb{E}[cS^2] - \sigma^2)^2 + c^2 Var(S^2)$$

$$= (c\underbrace{\mathbb{E}[S^2]}_{\sigma^2} - \sigma^2)^2 + c^2 \underbrace{Var(S^2)}_{\frac{2\sigma^4}{N-1}}$$

$$= (c\sigma^2 - \sigma^2)^2 + \frac{c^2 2\sigma^4}{N-1}$$

$$= \sigma^4(c-1)^2 + \frac{2c^2\sigma^4}{N-1} = MSE(cS^2)$$

$$\frac{d\,MSE}{dc} = \cancel{2}\cancel{\sigma^4}(c-1) + \frac{\cancel{4}^{2} c \cancel{\sigma^4}}{N-1} = 0$$

$$\Rightarrow c-1 + 2c/_{N-1} = 0$$

$$\Rightarrow (c-1)(N-1) + 2c = 0$$

$$\Rightarrow c(N-1) - (N-1) + 2c = 0$$

$$\Rightarrow c(N-1+2) = N-1$$

$$\Rightarrow c^* = \frac{N-1}{N+1}$$

Looking at cS^2

$$c^* S^2 = \frac{N-1}{N+1} \frac{1}{N-1} \sum_{n=1}^{N} (X_n - \bar{X})^2 = \frac{1}{N+1} \sum_{n=1}^{N} (X_n - \bar{X})^2$$