We've talked about

1. MoM
2. MLE

How do we evaluate/compare estimators?

---

**Defn: Mean-Squared Error (MSE)**

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

let $\hat{\theta}$ be an estimate of $\theta$.

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

$$MSE_\theta(\hat{\theta}) = \mathbb{E}[\underbrace{(\hat{\theta}-\theta)^2}_{\text{avg. squared deviation of } \hat{\theta} \text{ from } \theta}].$$

MSE depends on true (but unknown) value of $\theta$

avg. squared deviation of $\hat{\theta}$ from $\theta$.

---

If our estimator $\hat{\theta}$ is "good" the $MSE_\theta(\hat{\theta})$ is small.

Otherwise, MSE is large.

Idea: If I have two estimators $\hat{\theta}_1$ and $\hat{\theta}_2$

then we could say $\hat{\theta}_1$ is "better" than $\hat{\theta}_2$

if

$$MSE(\hat{\theta}_1) < MSE(\hat{\theta}_2).$$

---

**Defn: Bias**

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

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

We say that $\hat{\theta}$ is unbiased for $\theta$ if $B(\hat{\theta}) = 0$.

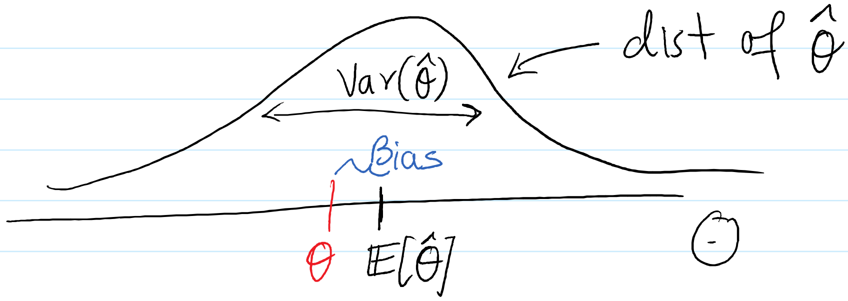
i.e. $\mathbb{E}[\hat{\theta}] = \theta$.

**Defn: Variance** Recall $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$.

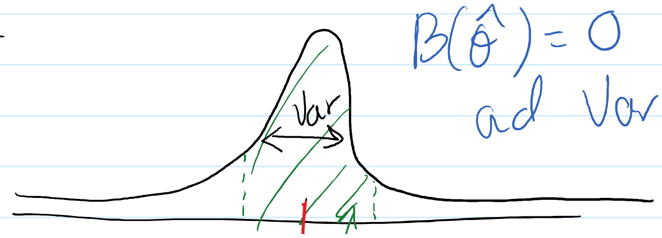
The variance of $\hat{\theta}$ is $Var(\hat{\theta})$.

---

Ex.

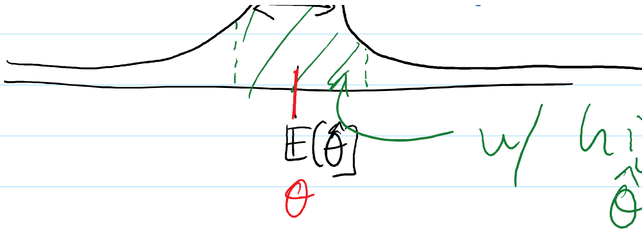



Ideally:




$B(\hat{\theta}) = 0$

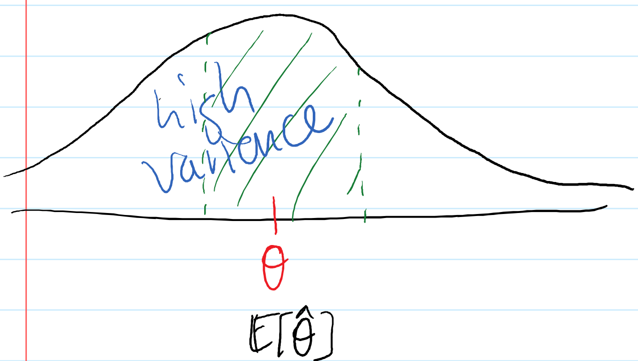
and $Var(\hat{\theta})$ small.




w/ high. prob. $\hat{\theta}$ is close to $\theta$

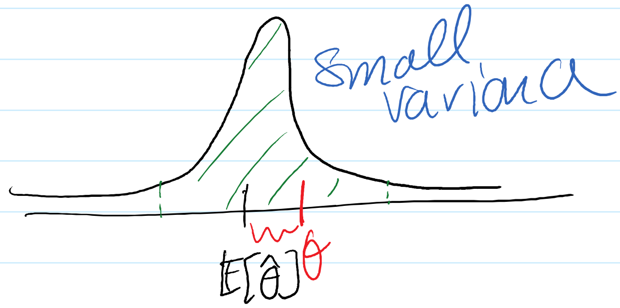
Not always true that unbiasedness is the "best"

<u>Unbiased $\hat{\theta}$</u>




lower prob. of $\hat{\theta}$ being close to $\theta$

<u>small bias</u>




higher prob. of $\hat{\theta}$ being close to $\theta$

---

**Theorem:** MSE = Bias$^2$ + Variance.

$$\underbrace{MSE(\hat{\theta})}_{\text{squared scale}} = \underbrace{B(\hat{\theta})}_{\text{un-squared scale}}{}^2 + \underbrace{Var(\hat{\theta})}_{\text{squared scale}}$$

<u>pf.</u>

$\theta$

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

$$= \mathbb{E}\Big[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2 + (\mathbb{E}[\hat{\theta}]-\theta)^2 + 2(\hat{\theta}-\mathbb{E}[\hat{\theta}])\underbrace{(\mathbb{E}[\hat{\theta}]-\theta)}_{\text{not random}}\Big]$$

$$= \mathbb{E}[(\hat{\theta}-\mathbb{E}[\hat{\theta}])^2] + \mathbb{E}[\underbrace{(\mathbb{E}[\hat{\theta}]-\theta)}_{\text{not random}}{}^2] + 2(\mathbb{E}[\hat{\theta}]-\theta)\underbrace{\mathbb{E}[\hat{\theta}-\mathbb{E}[\hat{\theta}]]}_{0}$$

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

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

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

---

<u>Ex.</u> $X_n$: $f$ ~~and~~ iid $\quad \mu = \mathbb{E}[X_n]$

$\sigma^2 = \mathrm{Var}(X_n)$

Consider $\hat{\mu} = \bar{X}$.

① $\mathbb{E}[\hat{\mu}] = \mu$ so $\hat{\mu}$ is unbiased for $\mu$

② $\mathrm{Var}(\hat{\mu}) = \sigma^2/N$

So $\text{MSE}(\hat{\mu}) = B(\hat{\mu})^2 + \text{Var}(\hat{\mu})$

$$= 0^2 + \sigma^2/N = \sigma^2/N$$

---

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

we know that

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

So $S^2$ is unbiased for $\sigma^2$.

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

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

Facts:

$Z \sim \chi^2(k)$

$\mathbb{E}[Z] = k$

$\text{Var}(Z) = 2k$

Certainly $\mathbb{E}\left[\frac{N-1}{\sigma^2} S^2\right] = (N-1)$

hence $\mathbb{E}[S^2] = \sigma^2$

and

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

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

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

thus

$$MSE(S^2) = B(S^2)^2 + Var(S^2)$$
$$= 0^2 + \frac{2\sigma^4}{N-1} = \boxed{\frac{2\sigma^4}{N-1}}$$

<u>Consider</u> $\hat{\sigma}^2_{MLE} = \frac{1}{N}\sum_{n=1}^{N}(X_n - \bar{X})^2 = \frac{N-1}{N} S^2$

$$MSE(\hat{\sigma}^2_{MLE}) = B(\hat{\sigma}^2_{MLE})^2 + Var(\hat{\sigma}^2_{MLE})$$

$$B(\hat{\sigma}^2_{MLE}) = B\left(\frac{N-1}{N}S^2\right) = E\left[\frac{N-1}{N}S^2\right] - \sigma^2$$
$$= \frac{N-1}{N}E[S^2] - \sigma^2$$
$$= \frac{N-1}{N}\sigma^2 - \sigma^2$$
$$= \frac{-1}{N}\sigma^2$$

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

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

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

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

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

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

which is smaller?

$$MSE(\hat{\sigma}^2_{MLE}) = \frac{2N-1}{N^2}\sigma^4 = \overbrace{\underbrace{\frac{2N-1}{N^2}\left(\frac{N-1}{2}\right)}_{<1\,?}}^{<1}\underbrace{\left(\frac{2}{N-1}\right)\sigma^4}_{MSE(s^2)}$$

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

$$\boxed{MSE(\hat{\sigma}^2_{MLE}) < MSE(S^2)}$$

---

<u>General Question:</u>

Is there some $c \in \mathbb{R}$ that minimizes MSE of

$$cS^2$$

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

$$\left\{ c = \frac{N-1}{N} \Rightarrow \hat{\sigma}^2_{MLE} \right.$$

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

$$= \left(cE[S^2] - \sigma^2\right)^2 + c^2 Var(S^2)$$

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

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

To minimize:

$$0 = \frac{\partial}{\partial c} MSE(cS^2) = \left[2(c-1) + \frac{4c}{N-1}\right]\sigma^4$$

$$\Rightarrow \quad 2(c-1) + \frac{4c}{N-1} = 0$$

$$\Rightarrow \quad 2c - 2 + \frac{4c}{N-1} = 0$$

$$\Rightarrow \quad \cancel{2}(N-1)c - \cancel{2}(N-1) + \overset{2}{\cancel{4}}c = 0$$

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

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

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

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

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

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

---

Want to find a "best" estimator of $\theta$.

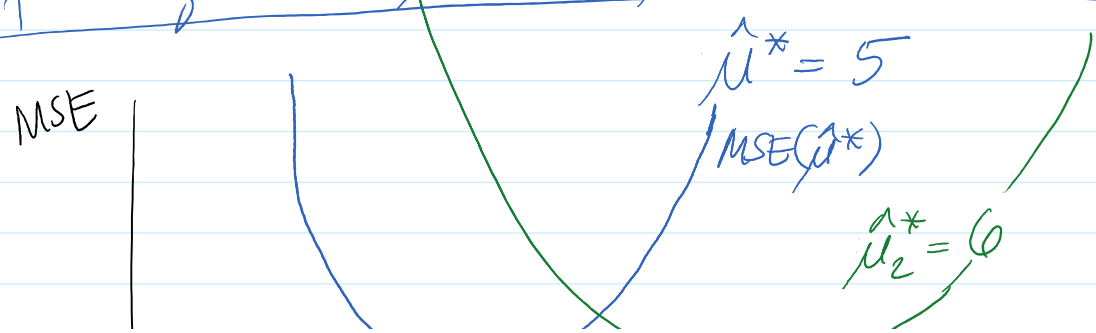
Fact! In general (if I'm too permissive) no globally "best estimator" exists.

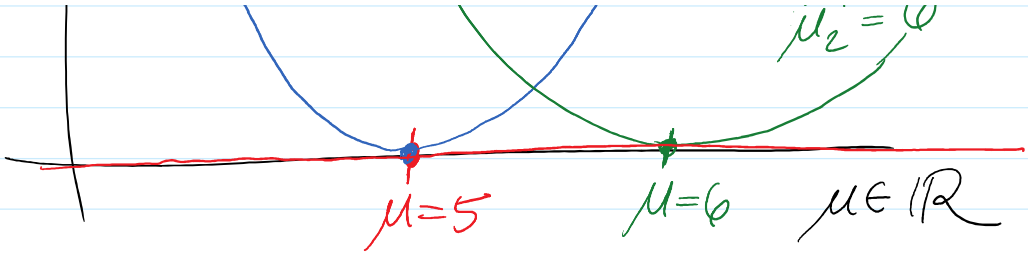
---

Ex. $X_n \overset{iid}{\sim} N(\mu, 1)$

Want to estimate $\mu$.

Want $\hat{\mu}^*$ so that $MSE_\mu(\hat{\mu}^*) \leq MSE_\mu(\hat{\mu})$ for any other $\hat{\mu}$ all $\mu \in \mathbb{R}$






$MSE(\hat{\mu}^*) = 0$ when $\mu = 5$

If I allow dumb estimators — I can't beat them at all possible values of $\mu$.
(↑ have a smaller MSE)

(unless of course my estimator is always perfectly correct)

Solution: restrict class of estimators to a "sensible" class.

Consider: unbiased estimator.

---

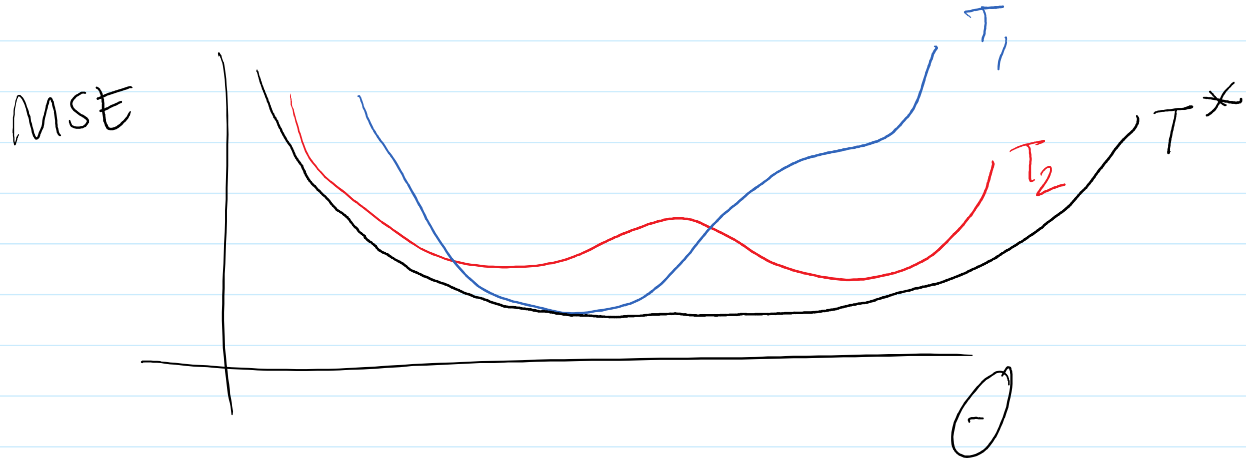
Defn: Best Unbiased Estimator

We call $T^*$ the best unbiased estimator of $\tau(\theta)$ — some fn of $\theta$ —
$\tau: \mathbb{R} \to \mathbb{R}$

if (1) $T^*$ is unbiased for $\tau(\theta)$
$$\mathbb{E}[T^*] = \tau(\theta)$$

② If $T$ is another unbiased est. for $\tau(\theta)$

then $MSE_\theta(T^*) \le MSE_\theta(T)$ for all $\theta \in \Theta$.




Notice: $B(T) = 0$ then $\boxed{MSE(T) = Var(T)}$

So another name for $T^*$ is the

uniformly minimum variance unbiased estimator

or UMVUE.