Lecture 7: Evaluation

Defn: Mean-Squared Error (MSE)

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

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

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

avg. sq. dist. of $\hat{\theta}$ to θ

If $\hat{\theta}$ is a good est. then $MSE(\hat{\theta})$ is small

Conversely not a good est. if $MSE(\hat{\theta})$ is large

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

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

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

If $B(\hat{\theta}) > 0$ then on avg. we over-estimate θ

" $B(\hat{\theta}) < 0$ " under- "

If $B(\hat\theta)=0$ we say that $\hat\theta$ is <u>unbiased</u>

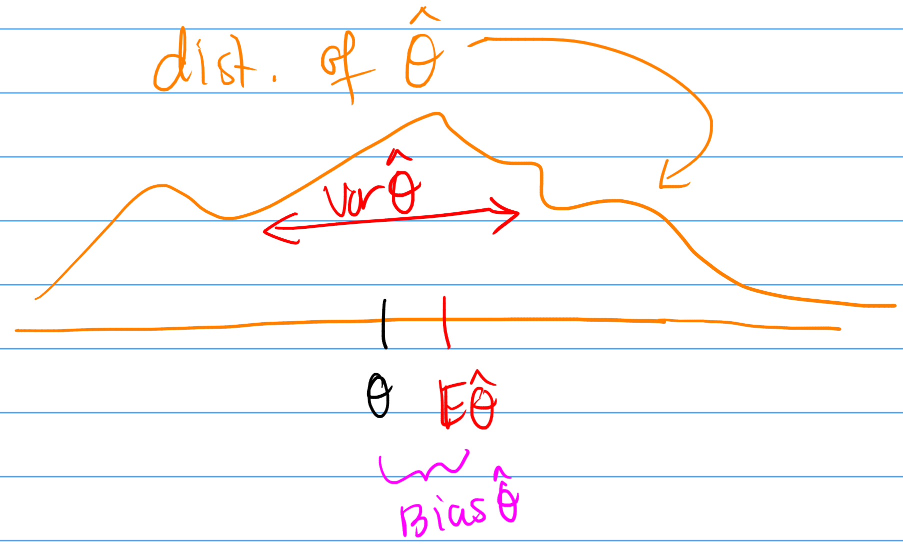
i.e. on avg. $\hat\theta$ correctly ests. θ.

Variance

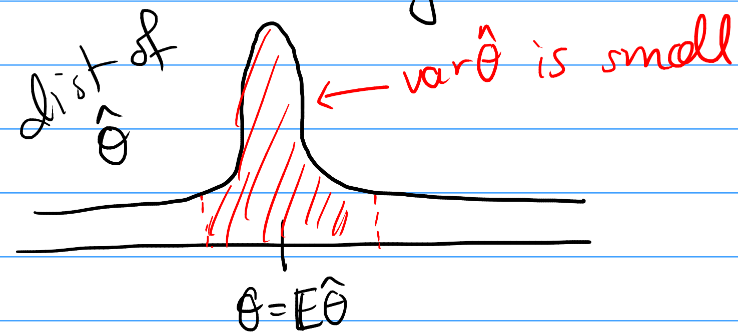
Recall that $\hat\theta$ is random so it has a variance

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

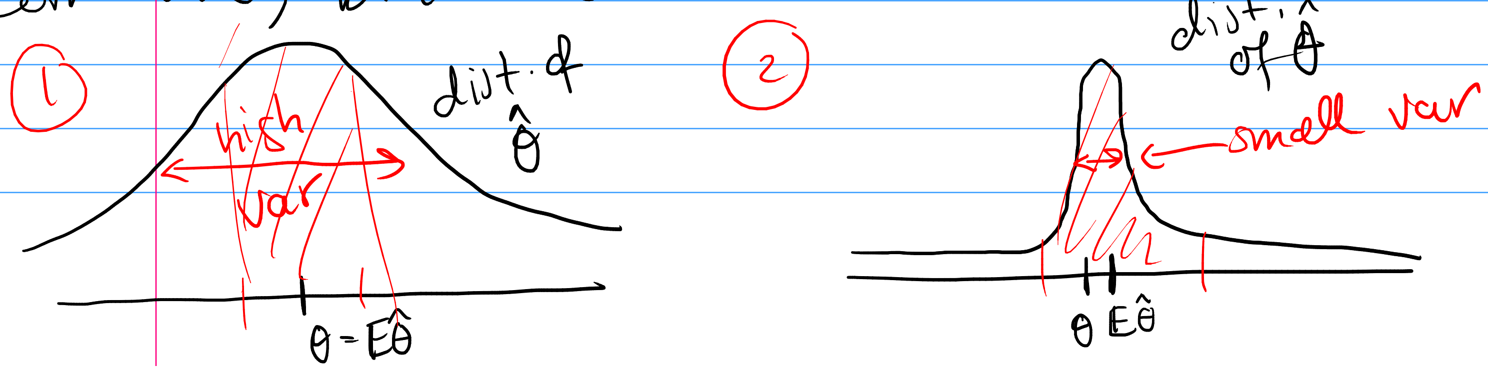
Ex.



Ideally $B(\hat\theta)$ is zero and my $Var\hat\theta$ is small



Sometimes, biased est. are better.



Theorem: MSE = bias^2 + Var

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

sq scale

not sq.

Pf.

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

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

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

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

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

$$= \mathrm{Var}(\hat\theta) + (\underbrace{\mathbb{E}\hat\theta - \theta}_{\text{Bias}})^2 + 2(\mathbb{E}\hat\theta - \theta)\underbrace{\mathbb{E}[\hat\theta - \mathbb{E}\hat\theta]}_{0}$$

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

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

Ex. Let $X_n \overset{iid}{\sim} f$

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

Consider $\hat\mu = \overline{X}$

We prev. showed that

(1) $\mathbb{E}\hat{\mu} = \mu$

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

So $$\begin{aligned} MSE(\hat{\mu}) &= B(\hat{\mu})^2 + \text{Var}(\hat{\mu}) \\ &= (\mathbb{E}(\hat{\mu}) - \mu)^2 + \sigma^2/N \\ &= \underbrace{(\mu - \mu)^2}_{0} + \sigma^2/N \\ &= \sigma^2/N \end{aligned}$$

$0 \rightarrow \hat{\mu}$ is unbiased for μ

Notice: if my est. is unbiased then MSE = Var

Ex. Consider $S_{N-1}^2 = \frac{1}{N-1}\sum_n (X_n - \bar{X})^2$

as an est. of σ^2.

$B(S^2) = \mathbb{E}S^2 - \sigma^2 = \sigma^2 - \sigma^2 = 0$

We showed that $\mathbb{E}[S_{N-1}^2] = \sigma^2$

So S_{N-1}^2 is unbiased for σ^2.

If $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ the we had a theorem:

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

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

$\mathbb{E}Z = k$

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

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

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

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

The MLE of σ^2 in the normal case is

$$\hat{\sigma}^2 = \frac{1}{N}\sum_n (X_n - \bar{X})^2 = \frac{N-1}{N} S_{N-1}^2$$

Let's calc the MSE.

Bias: $B(\hat{\sigma}^2) = \mathbb{E}[\hat{\sigma}^2] - \sigma^2$

$$\mathbb{E}\hat{\sigma}^2 = \mathbb{E}\left[\frac{N-1}{N} S_{N-1}^2\right] = \frac{N-1}{N}\mathbb{E}[S_{N-1}^2] = \frac{N-1}{N}\sigma^2$$

So $$B(\hat{\sigma}^2) = \frac{N-1}{N}\sigma^2 - \sigma^2 = \boxed{-\frac{1}{N}\sigma^2}$$

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

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

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

Combine to get MSE

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

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

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

(with $\overbrace{\left(\frac{2}{N-1}\right)\sigma^4}^{MSE(S^2_{N-1})}$)

$$= \underbrace{\frac{2N-1}{N^2}\left(\frac{N-1}{2}\right)}_{>1?\ <1?} MSE(S^2_{N-1})$$

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

So $MSE(\hat{\sigma}^2) < MSE(S^2_{N-1})$.

More generally: is there some constant c that minimizes MSE cS^2_{N-1}

$$\left\{\begin{array}{l} c = 1 \Rightarrow S^2_{N-1} \\ c = \frac{N-1}{N} \Rightarrow \hat{\sigma}^2 \end{array}\right.$$

$$\text{MSE}(cS^2)$$

$$= B(cS^2)^2 + \text{Var}(cS^2)$$

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

$$= \left[c\,\mathbb{E}[S^2] - \sigma^2\right]^2 + c^2\frac{2\sigma^4}{N-1}$$

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

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

Let's take a deriv.

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

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

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

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

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

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

$$c^* s^2 = \cancel{\frac{N-1}{N+1}} \frac{1}{\cancel{N-1}} \sum_n (X_n - \bar{X})^2$$

$$\boxed{= \frac{1}{N+1} \sum_n (X_n - \bar{X})^2}$$ ← minimizes MSE

EXAM 1

I want to find "best" estimator.

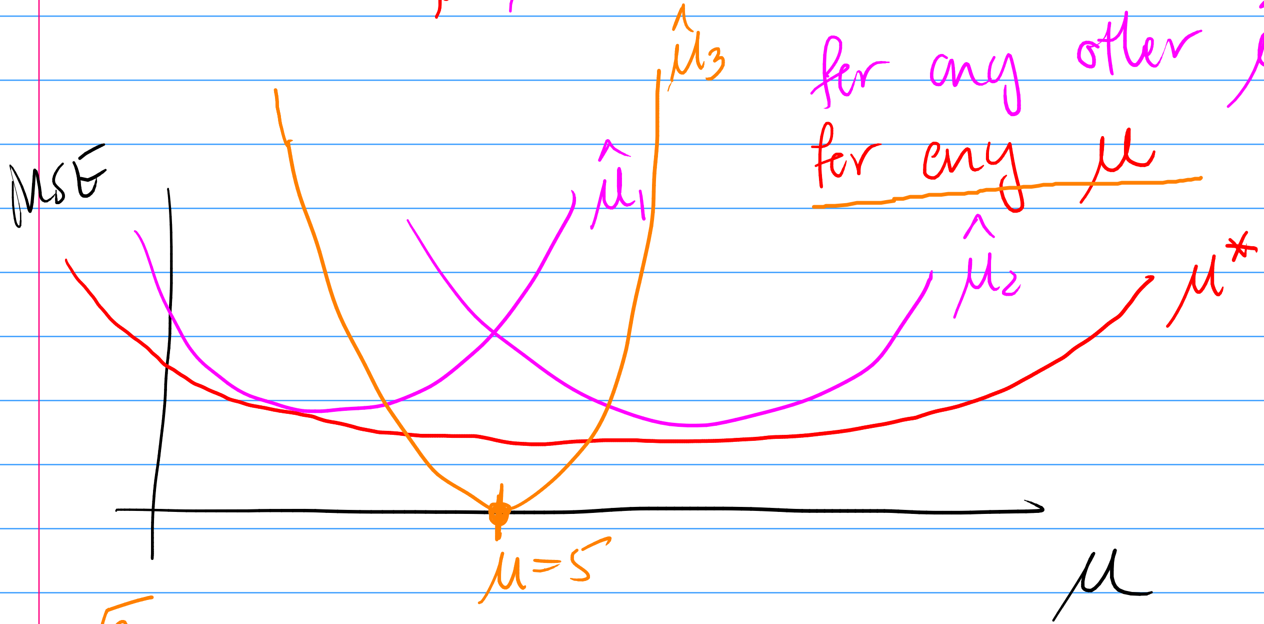
Problem: If I'm too permissive in what I allow to be an estimator then there is no answer to this question.

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

Want to find some μ^* that is the best estimator of μ, so that

$$MSE_\mu(\mu^*) \leq MSE_\mu(\hat{\mu})$$

for any other $\hat{\mu}$
for any μ



$\hat{\mu}_3 = 5$

We need to restrict allowable estimators.

One way: restrict to the class of unbiased estimators.