Lecture 7: 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.

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

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

If $\hat{\Theta}$ is good, MSE is small, vice-versa.

Idea: If I have two ests. $\hat{\theta}_1$ and $\hat{\theta}_2$ I might prefer the est. 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 $B(\hat{\theta}) > 0$ I tend to over-estimate
< 0 " under-estimate

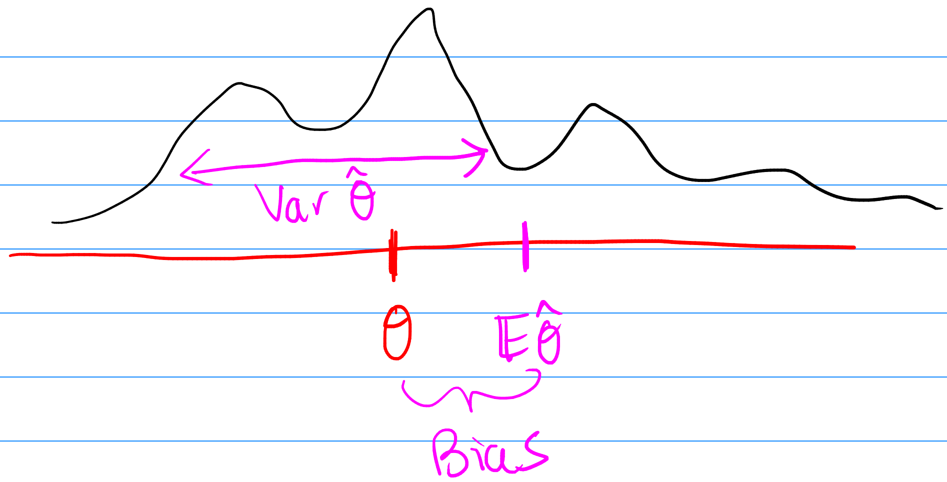
If $B(\hat{\theta}) = 0$ we say $\hat{\theta}$ is unbiased
i.e. gets pred. correct on avg.

Variance

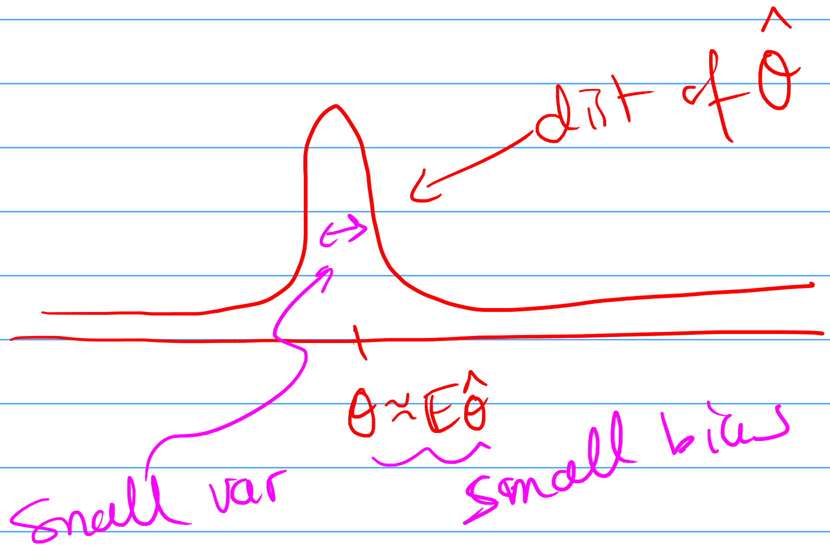
Recall that $\hat{\theta} = \hat{\theta}(\underset{\sim}{X})$ so it is random.

So $\hat{\theta}$ has a variance: $Var_\theta(\hat{\theta})$

Ex.



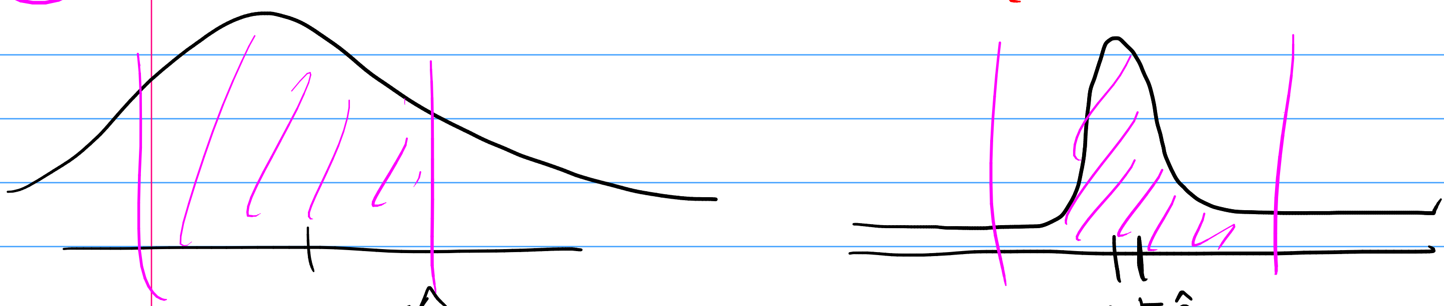
Ideally $B(\hat{\theta})$ is small and so is $Var(\hat{\theta})$



Sometimes, a (non-zero) biased est is better.

1 high var

2 low var



Theorem: $MSE = \text{bias}^2 + \text{Var}$

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

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

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

$$= \mathbb{E}[a^2] + \mathbb{E}[b^2] + 2\mathbb{E}[ab]$$

$\text{Var}(z)$

$\mathbb{E}[(z - \mathbb{E}z)^2]$

$$= \underbrace{\mathbb{E}[(\hat{\theta} - \mathbb{E}\hat{\theta})^2]}_{1} + \underbrace{\mathbb{E}[(\theta - \mathbb{E}\hat{\theta})^2]}_{3}$$

$$+ 2\underbrace{\mathbb{E}[(\hat{\theta} - \mathbb{E}\hat{\theta})(\mathbb{E}\hat{\theta} - \theta)]}_{3}$$

① $\mathbb{E}[(\hat{\theta}-\mathbb{E}\hat{\theta})^2] = \mathrm{Var}(\hat{\theta})$

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

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

$\mathbb{E}[aX] = a\mathbb{E}X$

③ $\mathbb{E}[(\theta-\mathbb{E}[\hat{\theta}])(\hat{\theta}-\mathbb{E}[\hat{\theta}])]$

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

$= (\mathbb{E}[\hat{\theta}]-\theta)(\underbrace{\mathbb{E}[\hat{\theta}]-\mathbb{E}[\hat{\theta}]}_{0})$

$= 0$

Ex. $X_n \overset{iid}{\sim} f$ where $\mu = \mathbb{E}(X_n)$

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

We prev. showed generally if $\hat{\mu} = \overline{X}$

then $\mathbb{E}[\hat{\mu}] = \mu$ ⇝ unbiased

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

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

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

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

Notice: If Bias $=0$ then MSE $=$ Var.

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

Showed that $E[S^2_{N-1}] = \sigma^2$

and so S^2_{N-1} is unbiased

Consider $X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ we also showed

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

So $Var\left[\frac{N-1}{\sigma^2} S^2_{N-1}\right] = 2(N-1)$

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

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

So since S^2_{N-1} was unbiased for σ^2 then

$$MSE = Var = \frac{2\sigma^4}{N-1}.$$

$$S_{N-1}^2 = \frac{1}{N-1}\sum_n (x_n - \bar{x})^2$$

The MLE of σ^2 in the $N(\mu, \sigma^2)$ case was

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

Let's get MSE.

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

So

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

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

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

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

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

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

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

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

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

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

Consider

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

So $$MSE(\hat{\sigma}^2) = \underbrace{\frac{(2N-1)(N-1)}{2N^2}}_{>1?\ <1?} MSE(S_{N-1}^2)$$

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

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

More generally what do I multiply S^2 by to get the lowest MSE?

$$MSE(cS^2)$$

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

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

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

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

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

take deriv wrt c set to zero, solve.

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

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

$$\Rightarrow cN - c - N + 1 + 2c = 0$$

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

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

So min MSE is at $c^* S^2 = \frac{N-1}{N+1} S^2 = \boxed{\frac{1}{N+1} \sum_n (X_n - \bar{X})^2}$

I want to find a "best" estimator.

<u>Problem</u>: If I am too permissive in what I call an estimator, there is no all around best option.

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

want to find some μ^* that is best so that

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

$\forall$ possible $\hat{\mu}$

$\forall \mu$