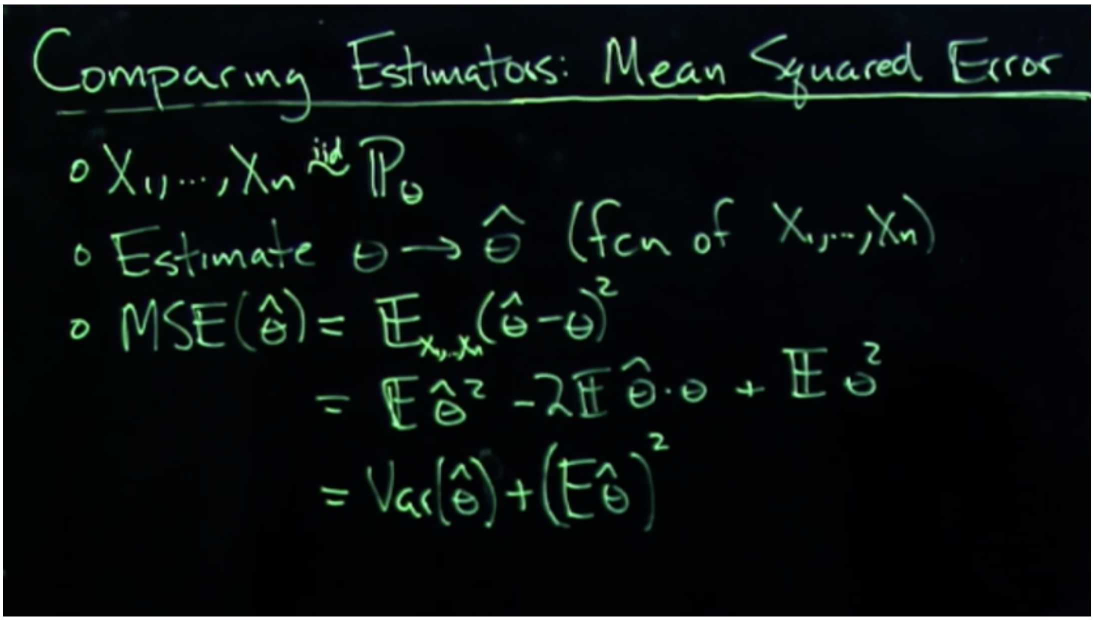
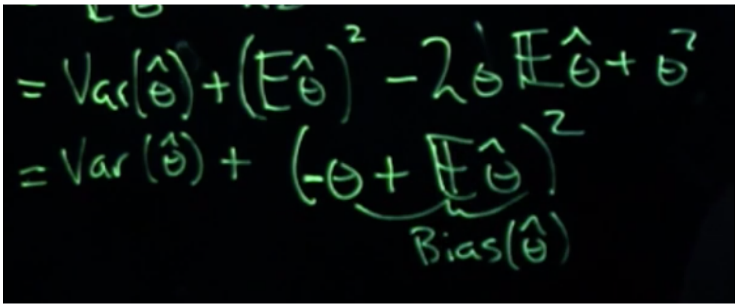
Unit 3 Recitations

Mean Squared Error

- observe X_1….X_n, rv’s distributed according to P_theta
- goal to estimate theta, using estimator theta_hat (our estimator)
- theta_hat is a function of X_1,,,X_n
- want to know how well this estimator performs - use MSE



- only theta_hat is a random variable here
- last line comes from knowing that the variance comes from var = first moment - (second moment)^2



A lot of the time we work with data where the expectation of the estimator is equal to the unknown parameter that we are trying to calculate - so the bias is 0 and we have that the mean squared error = variance(theta_hat)

- can get MSE lower than this variance using something called a shrinkage estimator (later)

1. Calculate MSE for some estimators

Poisson

① $X_1, \dots, X_n \overset{iid}{\sim} \text{Pois}(\lambda)$, $\hat{\lambda}_1 = X_1$
$\hat{\lambda}_2 = \overline{X}_n$

ⓐ
- $\text{Var}(\hat{\lambda}_1) = \text{Var}(X_1) = \lambda$
- $\text{Bias}(\hat{\lambda}_1) = \mathbb{E}\hat{\lambda}_1 - \lambda = \mathbb{E}X_1 - \lambda = 0$

$$\text{MSE}(\hat{\lambda}_1) = \text{Var}(\hat{\lambda}_1) + (\text{Bias}(\hat{\lambda}_1))^2$$
$$= \lambda + 0^2 = \lambda$$

a) bias is zero because expectation of a lambda_hat is lambda

ⓑ
- $\text{Var}(\hat{\lambda}_2) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)$

$$= \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}(X_i)$$
$$= \frac{\lambda}{n}$$

- $\text{Bias}(\hat{\lambda}_2) = \mathbb{E}\hat{\lambda}_2 - \lambda = \left(\mathbb{E}\frac{1}{n}\sum_{i=1}^{n} X_i\right) - \lambda$

$$= \lambda - \lambda = 0$$

b) step 1 using sample mean
variance of each X_i is lambda so it’s a ’n’ sum of lambdas
another unbiased estimator

- $\text{MSE}(\hat{\lambda}_2) = \text{Var}(\hat{\lambda}_2) + (\text{Bias}(\hat{\lambda}_2))^2$

$$= \frac{\lambda}{n} + 0^2 = \frac{\lambda}{n}$$

the MSE for this one is smaller than the previous by a factor of dividing by n, so it is a better estimator (since we’re using many X_i instead of just X_1)

Uniform

① $X_1,\dots,X_n \overset{iid}{\sim} \text{Unif}([0,\theta])$ $\qquad \hat{\Theta}_1 := 2\bar{X}_n$

$\hat{\Theta}_2 = \max_i X_i$

(a) • $\text{Var}(\hat{\Theta}_1) = \text{Var}\left(2\cdot\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{4}{n^2}\sum_{i=1}^{n}\text{Var}(X_i)$

$$= \frac{4}{n^2}\cdot\frac{n\cdot\theta^2}{12} = \frac{\theta^2}{3n}$$

• $\text{Bias}(\hat{\Theta}_1) = \mathbb{E}\hat{\Theta}_1 - \theta = \frac{1}{n}\sum_{i=1}^{n} 2\mathbb{E}X_i - \theta$

$$= 0$$

• $\text{MSE}(\hat{\Theta}_1) = \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n}$

- variance of uniform is sigma^2 /12
- the 4/n^2 comes from pulling out 2/n of the variance
- mean of this uniform is theta/2 which cancels with the

(b) • cdf of $\hat{\Theta}_2$: $F(x) = P(\hat{\Theta}_2 \le x)$

$$= P(\max_i X_i \le x)$$

$$= (P(X_1 \le x))^n$$

$$= \left(\frac{x}{\theta}\right)^n, \quad x \in [0,\theta]$$

• $f(x) = \frac{nx^{n-1}}{\theta^n}$

• $\mathbb{E}(\hat{\Theta}_2) = \int_0^\theta x\cdot\frac{nx^{n-1}}{\theta^n}\,dx = \frac{n}{n+1}\cdot\frac{x^{n+1}}{\theta^n}\Big|_0^\theta$

$$= \frac{n}{n+1}\theta$$

- X_i changes to X_1 because they are all equivalent
- bias is not 0

$$\mathbb{E}\hat{\Theta}_2^2 = \int_0^\theta x^2 \cdot \frac{nx^{n-1}}{\theta^n}\,dx = \frac{n}{n+2}\cdot \frac{x^{n+2}}{\theta^n}\Big|_0^\theta$$

$$= \frac{n}{n+2}\cdot \theta^2$$

- second moment, now we can calculate MSE

$$\circ\ \mathrm{MSE}(\hat{\Theta}_2) = \mathrm{Var}(\hat{\Theta}_2) + \left(\mathrm{Bias}(\hat{\Theta}_2)\right)^2$$

$$= \left(\frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2\right) + \left(\frac{n}{n+1}\theta - \theta\right)^2$$

$$= \theta^2\left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2} + \frac{1^2}{(n+1)^2}\right]$$

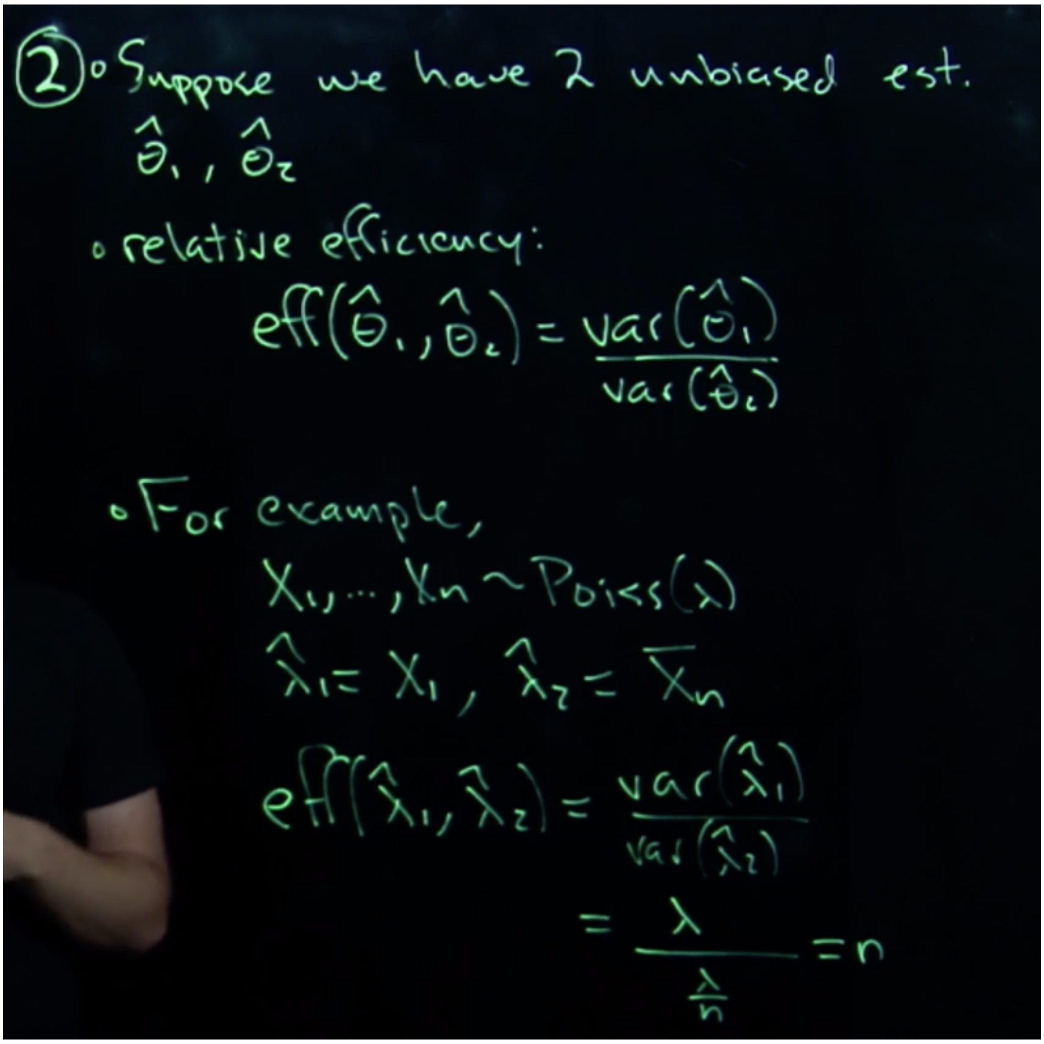
$$= \theta^2\left[\frac{n(n+1)^2 - n^2(n+2) + (n+2)}{(n+2)(n+1)^2}\right]$$

$$= \theta^2\left[\frac{n \quad + n+2}{(n+2)(n+1)^2}\right]$$

$$= \frac{2\theta^2}{(n+2)(n+1)}$$

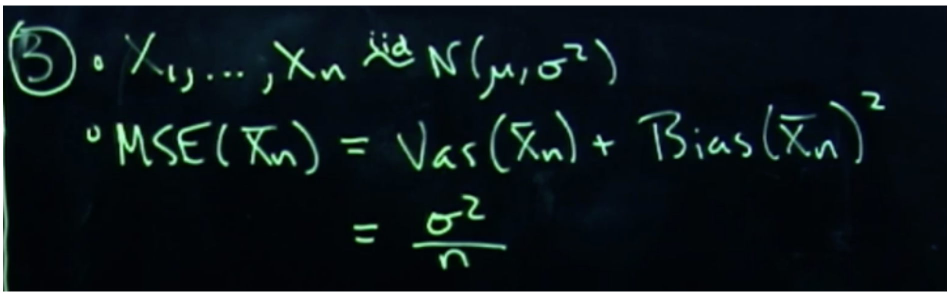
- the second MSE will perform better as it is dividing by a factor of n^2 instead of just n

2. Compare MSE of estimators

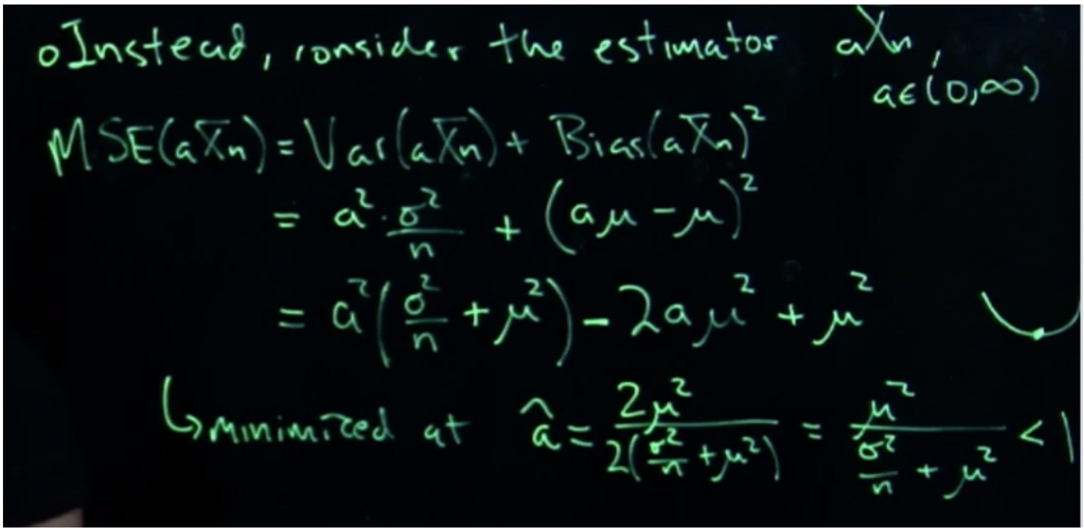


- var/ var because there's no bias on either
- the relative efficiency shows that the second MSE is more efficient by a factor of n as shown earlier

3. Shrinkage estimators

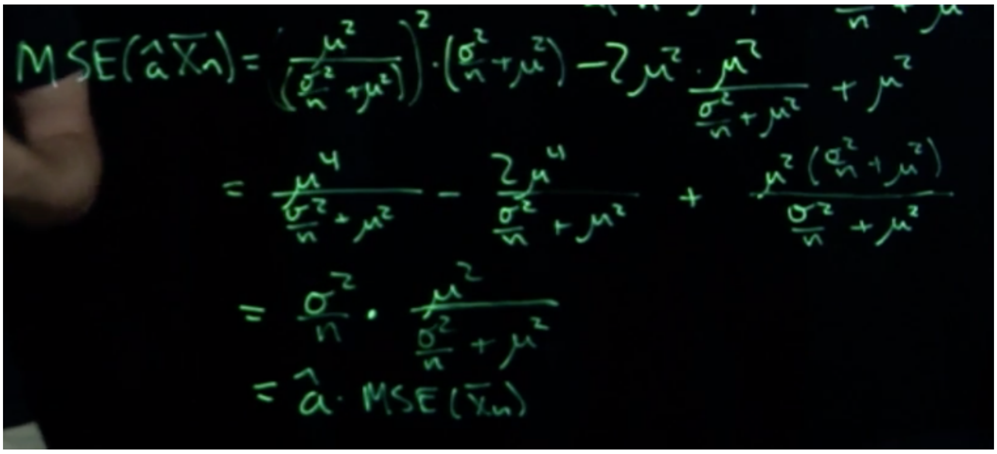


- baseline MSE



- this MSE is smaller

plug in a_hat for a



we’re decreasing the variance by adding a bit of bias which leads to a smaller MSE