

Unit 3 Recitations

Mean Squared Error

- observe X_1, \dots, X_n , rv's distributed according to P_θ
- goal to estimate θ , using estimator $\hat{\theta}$ (our estimator)
- $\hat{\theta}$ is a function of X_1, \dots, X_n
- want to know how well this estimator performs - use MSE

Comparing Estimators: Mean Squared Error

- $X_1, \dots, X_n \stackrel{iid}{\sim} P_\theta$
- Estimate $\theta \rightarrow \hat{\theta}$ (fcn of X_1, \dots, X_n)
- $MSE(\hat{\theta}) = \mathbb{E}_{X_1, X_n} (\hat{\theta} - \theta)^2$
 $= \mathbb{E} \hat{\theta}^2 - 2\mathbb{E} \hat{\theta} \cdot \theta + \mathbb{E} \theta^2$
 $= \text{Var}(\hat{\theta}) + (\mathbb{E} \hat{\theta})^2$

- only $\hat{\theta}$ is a random variable here
- last line comes from knowing that the variance comes from $\text{var} = \text{first moment} - (\text{second moment})^2$

$$\begin{aligned} &= \text{Var}(\hat{\theta}) + (\mathbb{E} \hat{\theta})^2 - 2\theta \mathbb{E} \hat{\theta} + \theta^2 \\ &= \text{Var}(\hat{\theta}) + \underbrace{(-\theta + \mathbb{E} \hat{\theta})^2}_{\text{Bias}(\hat{\theta})} \end{aligned}$$

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($\hat{\theta}$)

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

1. Calculate MSE for some estimators

Poisson

$$\textcircled{1} X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda), \quad \hat{\lambda}_1 = X_1 \\ \hat{\lambda}_2 = \bar{X}_n$$

$$\textcircled{a} \begin{aligned} &\circ \text{Var}(\hat{\lambda}_1) = \text{Var}(X_1) = \lambda \\ &\circ \text{Bias}(\hat{\lambda}_1) = \mathbb{E}\hat{\lambda}_1 - \lambda = \mathbb{E}X_1 - \lambda = 0 \\ &\text{MSE}(\hat{\lambda}_1) = \text{Var}(\hat{\lambda}_1) + (\text{Bias}(\hat{\lambda}_1))^2 \\ &\quad = \lambda + 0^2 = \lambda \end{aligned}$$

a) bias is zero because expectation of a λ_{hat} is λ

$$\textcircled{b} \begin{aligned} &\circ \text{Var}(\hat{\lambda}_2) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \\ &\quad = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &\quad = \frac{\lambda}{n} \\ &\circ \text{Bias}(\hat{\lambda}_2) = \mathbb{E}\hat{\lambda}_2 - \lambda = \left(\mathbb{E} \frac{1}{n} \sum_{i=1}^n X_i\right) - \lambda \\ &\quad = \lambda - \lambda = 0 \end{aligned}$$

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

$$\begin{aligned} \circ \text{MSE}(\hat{\lambda}_2) &= \text{Var}(\hat{\lambda}_2) + (\text{Bias}(\hat{\lambda}_2))^2 \\ &= \frac{\lambda}{n} + 0^2 = \frac{\lambda}{n} \end{aligned}$$

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

$$\textcircled{1} \quad X_1, \dots, X_n \stackrel{iid}{\sim} \text{Unif}([0, \theta]) \quad \hat{\theta}_1 = 2\bar{X}_n \\ \hat{\theta}_2 = \max_i X_i$$

$$\begin{aligned} \textcircled{a} \quad \bullet \text{Var}(\hat{\theta}_1) &= \text{Var}\left(2 \cdot \frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{4}{n^2} \sum_{i=1}^n \text{Var}(X_i) \\ &= \frac{4}{n^2} \cdot n \cdot \frac{\theta^2}{12} = \frac{\theta^2}{3n} \\ \bullet \text{Bias}(\hat{\theta}_1) &= \mathbb{E}\hat{\theta}_1 - \theta = \frac{1}{n} \sum_{i=1}^n \mathbb{E}X_i - \theta \\ &= 0 \\ \bullet \text{MSE}(\hat{\theta}_1) &= \frac{\theta^2}{3n} + 0^2 = \frac{\theta^2}{3n} \end{aligned}$$

- 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

$$\begin{aligned} \textcircled{b} \quad \bullet \text{CDF of } \hat{\theta}_2: \quad F(x) &= P(\hat{\theta}_2 \leq x) \\ &= P(\max_i X_i \leq x) \\ &= (P(X_1 \leq x))^n \\ &= \left(\frac{x}{\theta}\right)^n, \quad x \in [0, \theta] \\ \bullet f(x) &= \frac{n x^{n-1}}{\theta^n} \\ \bullet \mathbb{E}(\hat{\theta}_2) &= \int_0^\theta x \cdot \frac{n x^{n-1}}{\theta^n} dx = \frac{n}{n+1} \cdot \frac{x^{n+1}}{\theta^n} \Big|_0^\theta \\ &= \frac{n}{n+1} \theta \end{aligned}$$

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

$$\begin{aligned} E\hat{\Theta}_2^2 &= \int_0^{\Theta} x^2 \cdot \frac{nx^{n-1}}{\Theta^n} dx = \frac{n}{n+2} \cdot \frac{x^{n+2}}{\Theta^n} \Big|_0^{\Theta} \\ &= \frac{n}{n+2} \cdot \Theta^2 \end{aligned}$$

- second moment, now we can calculate MSE

$$\begin{aligned} \circ \text{MSE}(\hat{\Theta}_2) &= \text{Var}(\hat{\Theta}_2) + (\text{Bias}(\hat{\Theta}_2))^2 \\ &= \left(\frac{n}{n+2} \Theta^2 - \left(\frac{n}{n+1} \Theta \right)^2 \right) + \left(\frac{n}{n+1} \Theta - \Theta \right)^2 \\ &= \Theta^2 \left[\frac{n}{n+2} - \frac{n^2}{(n+1)^2} + \frac{1^2}{(n+1)^2} \right] \\ &= \Theta^2 \left[\frac{n(n+1)^2 - n^2(n+2) + (n+2)}{(n+2)(n+1)^2} \right] \\ &= \Theta^2 \left[\frac{n + n+2}{(n+2)(n+1)^2} \right] \\ &= \frac{2\Theta^2}{(n+2)(n+1)} \end{aligned}$$

- 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

② Suppose we have 2 unbiased est.
 $\hat{\theta}_1, \hat{\theta}_2$

• relative efficiency:

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{\text{var}(\hat{\theta}_1)}{\text{var}(\hat{\theta}_2)}$$

• For example,

$$X_1, \dots, X_n \sim \text{Pois}(\lambda)$$

$$\hat{\lambda}_1 = X_1, \hat{\lambda}_2 = \bar{X}_n$$

$$\begin{aligned} \text{eff}(\hat{\lambda}_1, \hat{\lambda}_2) &= \frac{\text{var}(\hat{\lambda}_1)}{\text{var}(\hat{\lambda}_2)} \\ &= \frac{\lambda}{\frac{\lambda}{n}} = n \end{aligned}$$

- 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

③ • $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$

$$\begin{aligned} \text{MSE}(\bar{X}_n) &= \text{Var}(\bar{X}_n) + \text{Bias}(\bar{X}_n)^2 \\ &= \frac{\sigma^2}{n} \end{aligned}$$

- baseline MSE

o Instead, consider the estimator $a\bar{X}_n$, $a \in (0, \infty)$

$$\begin{aligned} \text{MSE}(a\bar{X}_n) &= \text{Var}(a\bar{X}_n) + \text{Bias}(a\bar{X}_n)^2 \\ &= a^2 \cdot \frac{\sigma^2}{n} + (a\mu - \mu)^2 \\ &= a^2 \left(\frac{\sigma^2}{n} + \mu^2 \right) - 2a\mu^2 + \mu^2 \end{aligned}$$

↳ minimized at $\hat{a} = \frac{2\mu^2}{2(\frac{\sigma^2}{n} + \mu^2)} = \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2} < 1$

– this MSE is smaller
plug in \hat{a} for a

$$\begin{aligned} \text{MSE}(\hat{a}\bar{X}_n) &= \left(\frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2} \right)^2 \cdot \left(\frac{\sigma^2}{n} + \mu^2 \right) - 2\mu^2 \cdot \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2} + \mu^2 \\ &= \frac{\mu^4}{\frac{\sigma^2}{n} + \mu^2} - \frac{2\mu^4}{\frac{\sigma^2}{n} + \mu^2} + \frac{\mu^2(\frac{\sigma^2}{n} + \mu^2)}{\frac{\sigma^2}{n} + \mu^2} \\ &= \frac{\sigma^2}{n} \cdot \frac{\mu^2}{\frac{\sigma^2}{n} + \mu^2} \\ &= \hat{a} \cdot \text{MSE}(\bar{X}_n) \end{aligned}$$

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