Lecture 31: Comparing estimators

Ciaran Evans

Recap: MSE

Let $\widehat{\theta}$ be an estimator of θ. The **mean squared error** (MSE) of $\widehat{\theta}$ is

$$MSE(\widehat{\theta}) = \mathbb{E}_\theta[(\widehat{\theta} - \theta)^2] = Var(\widehat{\theta}) + Bias^2(\widehat{\theta})$$

Common approaches:

- Try to make MSE small
- Restrict to unbiased estimators; try to make variance small

MSE and consistent estimators

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

Def: If $\hat{\theta} \xrightarrow{P} \theta$, we say $\hat{\theta}$ is a consistent estimator of θ

Theorem: If $MSE(\hat{\theta}) \to 0$ as $n \to \infty$, then $\hat{\theta} \xrightarrow{P} \theta$

(i.e. if bias $\to 0$ and variance $\to 0$, then $\hat{\theta}$ is consistent)

Pf: WTS $\forall \varepsilon > 0$, $P(|\hat{\theta} - \theta| > \varepsilon) \to 0$ as $n \to \infty$

Let $\varepsilon > 0$

$$P(|\hat{\theta} - \theta| > \varepsilon) = P((\hat{\theta}-\theta)^2 > \varepsilon^2)$$

$$\leq \frac{\mathbb{E}[(\hat{\theta}-\theta)^2]}{\varepsilon^2} \quad \text{(Markov's)}$$

$$= \frac{MSE(\hat{\theta})}{\varepsilon^2} \to 0$$

//

MSE and consistent estimators

$\hat{\theta} \xrightarrow{P} \theta$ does ~~not~~ necessarily imply $MSE(\hat{\theta}) \to 0$

Example: $U \sim \text{Uniform}(0,1)$

$$X_n = \sqrt{n}\, \mathbb{1}\{U \le \tfrac{1}{n}\}$$

$X_n \xrightarrow{P} 0$ Pf: $P(|X_n| > \varepsilon) = P(U \le \frac{1}{n}) \to 0$ as $n \to \infty$

$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = 1 \not\to 0$$

$$X_n^2 = \begin{cases} 0 & \text{w/ prob } 1 - \frac{1}{n} \\ n & \text{w/ prob } \frac{1}{n} \end{cases}$$

$$\mathbb{E}[X_n^2] = 0(1 - \tfrac{1}{n}) + n(\tfrac{1}{n}) = 1$$

Issue: X_n is unbounded (as $n \to \infty$, $\sqrt{n} \to \infty$)

MSE example

Suppose $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Consider two estimates of σ^2:

$$\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \qquad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

Activity: Compute the MSE for $\widehat{\sigma}^2$ and s^2 (see handout).

$$\mathbb{E}[s^2] = \frac{1}{n-1}\mathbb{E}\left[\textstyle\sum_i (X_i - \overline{X})^2\right]$$

$$= \frac{\sigma^2}{n-1}\mathbb{E}\Big[\underbrace{\frac{1}{\sigma^2}\textstyle\sum_i (X_i - \overline{X})^2}_{\chi^2_{n-1}}\Big] = \frac{\sigma^2}{n-1}(n-1) = \sigma^2$$

$$\Rightarrow \text{Bias}(s^2) = 0$$

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}\left[\frac{n-1}{n}s^2\right] = \frac{n-1}{n}\mathbb{E}[s^2] = \frac{(n-1)}{n}\sigma^2$$

$$\Rightarrow \text{Bias}(\hat{\sigma}^2) = -\frac{\sigma^2}{n}$$

Intuition: $\mathbb{E}\left[\frac{1}{n}\sum_i (X_i - \mu)^2\right] = \sigma^2$

$\overline{X}$ minimizes $\sum_i (X_i - a)^2$ $\Rightarrow$ $\sum_i (X_i - \overline{X})^2 \leq \sum_i (X_i - \mu)^2$

$$\mathrm{Var}(S^2) = \left(\frac{1}{n-1}\right)^2 \mathrm{Var}\left(\sum_i (X_i - \overline{X})^2\right)$$

$$= \left(\frac{\sigma^2}{n-1}\right)^2 \mathrm{Var}\Big(\underbrace{\frac{1}{\sigma^2}\sum_i (X_i-\overline{X})^2}_{\chi^2_{n-1}}\Big) = \frac{\sigma^4 \cdot 2(n-1)}{(n-1)^2} = \frac{2\sigma^4}{n-1}$$

$$\mathrm{Var}(\hat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 \mathrm{Var}(S^2) = \frac{2\sigma^4(n-1)}{n^2}$$

$$\Rightarrow \mathrm{MSE}(S^2) = \mathrm{Bias}^2(S^2) + \mathrm{Var}(S^2) = 0 + \frac{2\sigma^4}{n-1}$$

$$\mathrm{MSE}(\hat{\sigma}^2) = \mathrm{Bias}^2(\hat{\sigma}^2) + \mathrm{Var}(\hat{\sigma}^2)$$

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

$$\mathrm{MSE}(\hat{\sigma}^2) < \mathrm{MSE}(S^2)$$

$$\frac{2n-1}{n^2} < \frac{2}{n-1} \iff \frac{n-1}{n^2} < \frac{2}{2n-1} \iff (n-1)(2n-1) < 2n^2$$

$$2n^2 - n - 2n + 1 < 2n^2$$

Best unbiased estimators

Suppose we restrict ourselves to **unbiased** estimators.

Definition (best unbiased estimator):

Cramer-Rao lower bound

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} Poisson(\lambda)$

Why MLEs are nice

Let θ be a parameter of interest, and $\widehat{\theta}$ be the maximum likelihood estimator from a sample of size n. Under regularity conditions, $\widehat{\theta}$ satisfies the following properties:

- $\widehat{\theta} \xrightarrow{p} \theta$
- $\sqrt{n}(\widehat{\theta} - \theta) \xrightarrow{d} N(0, \mathcal{I}_1^{-1}(\theta))$