Comparing estimators

Ciaran Evans

Course so far

- Maximum likelihood estimation
- Tools for asymptotic results
- Asymptotic properties of MLEs

Questions:

- How else could we estimate parameters?
- What makes a good estimator? How do we compare estimators?
- Why is maximum likelihood estimation so widespread?
- What happens to MLEs if our model is *wrong*?

Example

Suppose $X_1, \ldots, X_n \overset{iid}{\sim} Uniform(0, \theta)$. How could I estimate θ?

1. MLE: $\hat{\theta} = X_{(n)}$

2. $\mathbb{E}[X] = \frac{\theta}{2}$ $\quad\quad$ $\bar{X} \approx \mathbb{E}[X]$ $\Rightarrow \hat{\theta} = 2\bar{X}$

3. $\mathbb{E}[X^2] = \frac{\theta^2}{3} \Rightarrow \hat{\theta} = \sqrt{\frac{3}{n}\sum_{i=1}^{n} X_i^2}$

4. $\hat{\theta} = 5$ (probably a terrible estimate)

Example

Suppose $X_1, \dots, X_n \overset{iid}{\sim} Uniform(a, b)$. How could I estimate a and b?

① MLE: $\hat{a} = X_{(1)}, \quad \hat{b} = X_{(n)}$

② $\mathbb{E}[X] = \dfrac{a+b}{2} = \mu_1 \qquad \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$

$\mathbb{E}[X^2] = \frac{1}{3}(a^2 + ab + b^2) = \mu_2 \qquad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$$b = 2\mu_1 - a$$

$$\Rightarrow \mu_2 = \frac{1}{3}\left(a^2 + a(2\mu_1 - a) + (2\mu_1 - a)^2\right)$$

$$= \frac{1}{3}\left(a^2 - 2a\mu_1 + 4\mu_1^2\right)$$

$\vdots$ (algebra)

$a = \mu_1 - \sqrt{3(\mu_2 - \mu_1)^2} \qquad b = \mu_1 + \sqrt{3(\mu_2 - \mu_1)^2}$

$\hat{a} = \hat{\mu}_1 - \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1)^2} \qquad \hat{b} = \hat{\mu}_1 + \sqrt{3(\hat{\mu}_2 - \hat{\mu}_1)^2}$

Method of moments

Let $X_1, ..., X_n$ be a sample from a distribution with probability function $f(x|\theta_1, ..., \theta_k)$, with k parameters $\theta_1, ..., \theta_k$.

Let

$$\mu_1 = \mathbb{E}[X] = g_1(\theta_1, \dots, \theta_k) \qquad \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$\mu_2 = \mathbb{E}[X^2] = g_2(\theta_1, \dots, \theta_k) \qquad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

$$\vdots$$

$$\mu_k = \mathbb{E}[X^k] = g_k(\theta_1, \dots, \theta_k) \qquad \hat{\mu}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

WLLN: $\hat{\mu}_i$ close to μ_i for large n

The method of moments (MoM) approach estimates $\theta_1, \dots, \theta_k$ by solutions to

$$\hat{\mu}_1 = g_1(\hat{\theta}_1, \dots, \hat{\theta}_k)$$

$$\vdots$$

$$\hat{\mu}_k = g_k(\hat{\theta}_1, \dots, \hat{\theta}_k)$$

Example

Suppose $X_1, ..., X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. want: $\hat{\mu}$ and $\hat{\sigma}^2$

$\mu_1 = \mathbb{E}[X] = \mu$ $\qquad \hat{\mu}_1 = \frac{1}{n}\sum_{i=1}^{n} X_i$

$\mu_2 = \mathbb{E}[X^2] = \sigma^2 + \mu^2$ $\qquad \hat{\mu}_2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$

$\hat{\mu} = \bar{X}$ ✓

$\sigma^2 = \mu_2 - \mu^2 \Rightarrow \hat{\sigma}^2 = \hat{\mu}_2 - \hat{\mu}_1^2$

$= \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$

$= \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$ ✓

In this case, $\hat{\mu}_{MOM} = \hat{\mu}_{MLE} = \bar{X}$

$\hat{\sigma}^2_{MOM} = \hat{\sigma}^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2$

What makes a good estimator?

Suppose $X_1, ..., X_n \overset{iid}{\sim} Uniform(0, \theta)$. Two possible estimates:

$$\text{MLE: } \widehat{\theta} = X_{(n)} \qquad \text{MoM: } \widehat{\theta} = 2\overline{X}$$

Question: How would I choose between these estimators?

Nice properties:

consistency: $\hat{\theta} \xrightarrow{p} \theta$ as $n \to \infty$

(true here for both $\hat{\theta}_{MLE} = X_{(n)}$ and $\hat{\theta}_{MoM} = 2\overline{X}$)

Low variance: $Var(\hat{\theta})$

unbiased: $\mathbb{E}[\hat{\theta}] = \theta$

Bias, variance, and MSE (mean squared error)

MSE: Let $\hat{\theta}$ be an estimator of θ

The MSE of $\hat{\theta}$ is $E_\theta[(\hat{\theta}-\theta)^2]$

θ is the value of the parameter

$$E_\theta[(\hat{\theta}-\theta)^2] = Var(\hat{\theta}) + \underbrace{(E_\theta[\hat{\theta}]-\theta)^2}_{Bias^2}$$

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

From HW: if $MSE(\hat{\theta}) \to 0$, then $\hat{\theta} \xrightarrow{q.m.} \theta$ and so $\hat{\theta} \xrightarrow{P} \theta$

One approach to choosing $\hat{\theta}$: minimize MSE

Another approach: restrict to unbiased estimators, minimize variance

Example

Suppose $X_1, ..., X_n \stackrel{iid}{\sim} Uniform(0, \theta)$.