Slide 5: Estimation Methods

STATS 511: Statistical Inference

Kean Ming Tan

Outline

Here are a list of topics we are going to cover throughout the next few lectures:

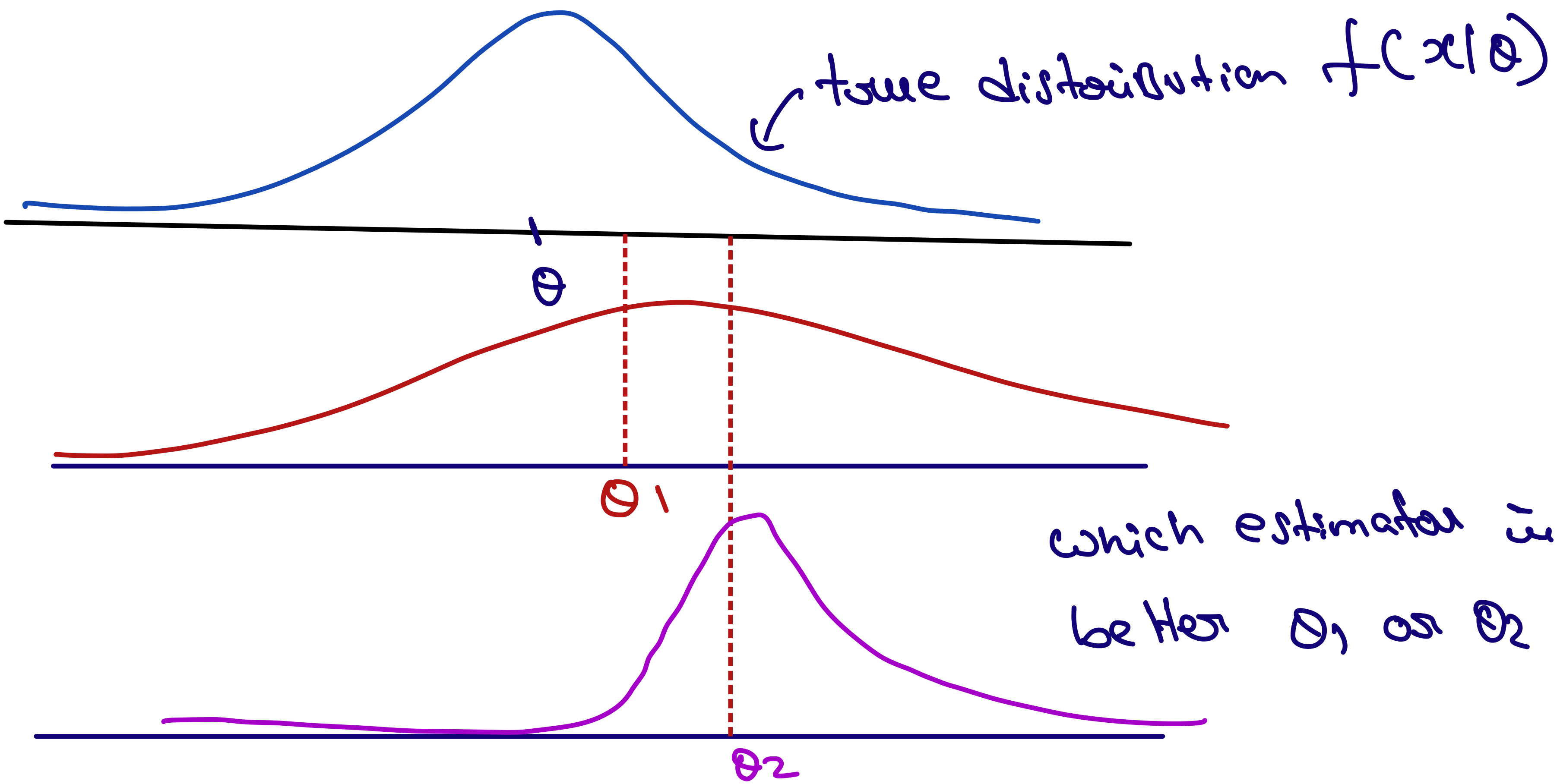
- Bias and Mean Squared Error
- Methods of Moments
- Generalized Method of Moment
- Maximum Likelihood Estimator

This slide set covers parts of Section 7.1–7.2 in the textbook

Reading: Section 7.1–7.2 in the textbook

Motivation

Suppose that you observe n data points $x_1, \ldots, x_n$ that are sampled from f_θ. Your goal is to use the data you observe an come up with an estimate for the unknown parameter θ. Let $\hat{\theta}_1$ be an estimate you come up with and let $\hat{\theta}_2$ be an estimate I come up with. How do we evaluate whether $\hat{\theta}_1$ or $\hat{\theta}_2$ is better?



Bias

Definition: An estimator $\hat{\theta}$ for θ is bias if $E(\hat{\theta}) \neq \theta$. The **bias** can be expressed by

$$Bias(\theta) = E_{\theta}(\hat{\theta}) - \theta.$$

If $Bias(\hat{\theta}) = 0$, then we say that $\hat{\theta}$ is an unbiased estimator of θ.

Question: Is Bias alone sufficient to evaluate whether an estimator is good or bad?

Unbiased Estimator - Key Notes

if $\mathbb{E}[\hat{\theta}] = \theta \implies$ Unbiased Estimator $\hat{\theta}$

① if we repeat the experiment many times, sometimes $\hat{\theta}$ will be higher and sometimes lower than θ, but on average, it will be correct.

② even if a single estimate $\hat{\theta}$ is far from θ, over many trails, the Estimator does not systematically overestimate or underestimate.

An unbiased estimator does not guarantee a good estimate for small samples, but over repeated experiments, it will be correct on average.

Example: Estimating Poisson Mean

Let $X_1, \dots, X_n \sim \mathrm{Poisson}(\lambda)$. Find an unbiased estimator for λ.

$\bar{X}$ is unbaised estimator of λ

$\mathbb{E}[\bar{X}] = \lambda$

S^2 is unbaised estimator of λ

$\mathbb{E}[S^2] = \lambda$

$a\bar{X} + (1-a)S^2 \implies$ unbaised estimator of λ

$\implies$ Bias alone cannot differentiate Estimator's.

Mean Squared Error

Definition: The mean squared error of an estimator $\hat{\theta}$ of a parameter θ is a function of θ defined by

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

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

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

$$+ 2\mathbb{E}\left[\left(\hat{\theta}-\underbrace{E[\hat{\theta}]}_{=0}\right)\left(E[\hat{\theta}]-\theta\right)\right]$$

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

Interpretation: The variance and bias go against each other. This is also referred to as bias-variance trade-off.

linear regression ⇒ unbaised Estimator.

ridge regression ⇒ Baised Estimator.

Example: Estimating Normal Variance

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. Here we want to estimate $\theta = \sigma^2$. Calculate the MSE of S^2. Is there a better estimator for σ^2 with smaller MSE?

$$S^2 = \frac{1}{n-1} \sum (x_i - \bar{x})^2$$

$$E[S^2] = \sigma^2$$

Consider $\tilde{\sigma}^2 = c \cdot \sum (x_i - \bar{x})^2$

Find c such that $MSE[\tilde{\sigma}^2]$ is minimized

FACT!

$$\frac{\sum (x_i - \bar{x})^2}{\sigma^2} \sim \chi^2_{n-1}$$

$$\mathbb{E}\left[\sum (x_i - \bar{x})^2\right] = \sigma^2 (n-1)$$

$$\mathrm{Var}\left(\sum (x_i - \bar{x})^2\right) = \sigma^4 \cdot 2 \cdot (n-1)$$

$$MSE[\tilde{\sigma}^2] = \text{var}(\tilde{\sigma}^2) + \text{Bias}^2(\tilde{\sigma}^2)$$

$$= c^2 \cdot 2\sigma^4(n-1) + \left(c\sigma^2(n-1) - \sigma^2\right)^2$$

$$= 2c^2\sigma^4(n-1) + c^2\sigma^4(n-1)^2 + \sigma^4 - 2c\sigma^4(n-1)$$

$$= \sigma^4\left(1 - 2c(n-1) + 2c^2(n-1) + c^2(n-1)^2\right)$$

$$\frac{d\,MSE[\tilde{\sigma}^2]}{dc} = \sigma^4\left(-2(n-1) + 4c(n-1) + 2c(n-1)^2\right) = 0$$

$$\Rightarrow \quad 2c(n-1)(2 + n - 1) = 2(n-1)$$

$$\Rightarrow \quad \boxed{c = n+1}$$

$\Rightarrow$ $\tilde{\sigma}^2 = \frac{1}{n+1}\sum(x_i - \bar{x})^2$ is better than

S^2 in term's of MSE

Method of Moments

Theorem: Assume that $X_1, \ldots, X_n$ form a random sample from a distribution that is indexed by a k-dimensional parameter θ and has at least k finite moments. Method of moments estimators are found by equating the first k sample moments to the corresponding k population methods, and solving the resulting simultaneous equation. That is, we set

$$\frac{1}{n}\sum_{i=1}^{n} X_i^j = E(X^j) \text{ for } j = 1, \ldots, k.$$

$$\frac{1}{n}\sum x_i = \mathbb{E}[x]$$

$$\frac{1}{n}\sum x_i^2 = \mathbb{E}[x^2]$$

$$\vdots$$

$$\frac{1}{n}\sum x_i^k = \mathbb{E}[x^k]$$

Example

Normal Distribution: Suppose that $X_1, \ldots, X_n$ are normally distributed with mean μ and variance σ^2. What are the method of moment estimators of (μ, σ^2)?

$$\frac{1}{n}\sum x_i = \mathbb{E}[X] = \mu \quad - (1)$$

$$\frac{1}{n}\sum x_i^2 = \mathbb{E}[X^2] = \mathrm{var}(X) + E[X]^2$$

$$= \sigma^2 + \mu^2 \quad - (2)$$

$$\Rightarrow \quad \tilde{\mu} = \bar{X}$$

$$\tilde{\sigma}^2 = \frac{1}{n}\sum x_i^2 - \bar{X}^2 = \frac{1}{n}\left(\sum x_i^2 - n\bar{X}^2\right)$$

$$= \frac{n-1}{n} S^2$$

$\Rightarrow$ $\hat{\mu} = \bar{X}$, $\hat{\sigma}^2 = \frac{n-1}{n} S^2$ are the MOM estimator for (μ, σ^2)

Generalized method's of moment's Estimator

① Pick $\Phi(x_i|\theta)$ s.t $\mathbb{E}[\Phi(x_i|\theta)]=0$

② Estimate θ by setting $\frac{1}{n}\sum_{i=1}^{n}\Phi(x_i|\theta)=0$

Ex: Linear regression

$Y_i = x_i^T\beta + \varepsilon_i$ $\quad x_i \perp\!\!\!\perp \varepsilon_i \quad \mathbb{E}[x_i\varepsilon_i]=0$

$\mathbb{E}[x_i\varepsilon_i]=0$

$\Rightarrow \mathbb{E}[x_i(Y_i - x_i^T\beta)]=0$

therefore Pick $\phi(x_i, y_i|\theta) = x_i(Y_i - x_i^T\beta)$ $(\theta=\beta)$

$$\Rightarrow \mathbb{E}[\phi(x_i, y_i|\theta)]=0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}\Phi(x_i, y_i|\theta)=0$$

$$\Rightarrow \frac{1}{n}\sum_{i=1}^{n}x_i(y_i - x_i^T\beta)=0$$

$$\Rightarrow \sum_{i=1}^{n}x_iy_i - \hat{\beta}\sum_{i=1}^{n}x_ix_i^T=0$$

$$\Rightarrow \hat{\beta} = \frac{\sum_{i=1}^{n}x_iy_i}{\sum_{i=1}x_ix_i^T}$$

Example: $\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{pmatrix} \right]$

Goal is to estimate $\sigma_x^2, \sigma_y^2, \rho$

$\mathbb{E}[x_i^2] = \sigma_x^2 \qquad \mathbb{E}[x_i y_i] = \rho\sigma_x\sigma_y$

$\mathbb{E}[y_i^2] = \sigma_y^2$

Construct $\phi(x_i, y_i | \theta) = x_i y_i - \rho\sigma_x\sigma_y$

$\phi(x_i|\theta) = x_i^2 - \sigma_x^2$

$\phi(y_i|\theta) = y_i^2 - \sigma_y^2$

$\Rightarrow$ using Generalized MoM

$$\frac{1}{n}\sum x_i y_i = \rho\sigma_x\sigma_y \qquad \tilde{\sigma}_x^2 = \frac{1}{n}\sum x_i^2$$

$$\frac{1}{n}\sum x_i^2 = \sigma_x^2 \quad \Longrightarrow \quad \tilde{\sigma}_y^2 = \frac{1}{n}\sum y_i^2$$

$$\frac{1}{n}\sum y_i^2 = \sigma_y^2 \qquad \tilde{\rho} = \frac{\frac{1}{n}\sum x_i y_i}{\tilde{\sigma}_x^2\,\tilde{\sigma}_y^2}$$

Up Next - Maximum Likelihood Estimator