Statistics

Point Estimation

Shiu-Sheng Chen

Department of Economics
National Taiwan University

Fall 2019

Section 1

Estimation

Statistical Inference

- Given a random sample with sample size n

$$\{X_1, X_2, \ldots, X_n\} \sim^{i.i.d.} f(x;\theta)$$

where θ is an unknown population parameter.

- For example, suppose we are interested in θ, which is the (unknown) population proportion of NTU students, who have a significant other.

$$\{X_1, X_2, \ldots, X_n\} \sim^{i.i.d.} \text{Bernoulli}(\theta)$$

Statistical Inference

- We would like to find a good guess of the parameter θ: a point estimator.
- That is, we would like to come up with a random variable

$$\hat{\theta} = \delta(X_1, X_2, \ldots, X_n)$$

that we expect to be close to θ.

Estimator

Definition

Let $\{X_1, X_2, \dots, X_n\}$ be a random sample from the joint distribution indexed by a parameter $\theta \in \Theta$. A function $\hat{\theta} = \delta(X_1, X_2, \dots, X_n)$ is called a point estimator of the parameter θ.

- Θ is called a parameter space.
- When $X_1 = x_1$, $X_2 = x_2$,...,$X_n = x_n$ are *observed*, then $\delta(x_1, x_2, \dots x_n)$ is called the point estimate of θ.
- Every estimator is also a statistic (by nature of being a function of a random sample).

Estimator

- There can be more than one unknown parameter:

$$\{X_i\}_{i=1}^{n} \sim^{i.i.d.} f(x, \theta_1, \theta_2, \ldots, \theta_k)$$

- For example, $(\theta_1,\ \theta_2) = (\mu,\ \sigma^2)$ denotes the (unknown) population mean and variance of the S&P500 stock returns.

$$\{X_1, X_2, \ldots, X_n\} \sim^{i.i.d.} N(\mu, \sigma^2)$$

- Estimators:

$$\hat{\theta}_1 = \hat{\mu} = \delta_1(X_1, X_2, \ldots, X_n)$$
$$\hat{\theta}_2 = \hat{\sigma}^2 = \delta_2(X_1, X_2, \ldots, X_n)$$

How to Guess?

- Analogy Principle (類比原則)
    - Method of Moments (動差法)
- Method of Maximum Likelihood (最大概似法)

Analogy Principle

- To estimate the population mean $\mu = E(X)$, use the sample mean

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- To estimate the population variance $\sigma^2 = Var(X) = E(X - E(X))^2$, use the sample variance

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2 \quad \text{or} \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$$

- In general, to estimate the population moments $m_j = E(X^r)$, use the sample moments

$$\frac{1}{n}\sum_{i=1}^{n} X_i^j$$

Analogy Principle

- To estimate distribution function $F_X(x) = P(X \leq x)$, use the empirical distribution function.

$$\hat{F}_n(x) = \frac{\sum_{i=1}^{n} I_{\{X_i \leq x\}}}{n}$$

Method of Moments

- The method of moments is simply an application of analogy principle.
- Suppose that

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} f(x, \theta_1, \theta_2, \ldots, \theta_k)$$

- j-th population moment, which is a function of unknown parameters

$$E(X^j) = m_j(\theta_1, \theta_2, \ldots, \theta_k)$$

Example

- For example, let $X \sim \text{Uniform}[\theta_1, \theta_2]$,

$$E(X) = m_1(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} x \frac{1}{\theta_2 - \theta_1} dx = \frac{\theta_1 + \theta_2}{2}$$

$$E(X^2) = m_2(\theta_1, \theta_2) = \int_{\theta_1}^{\theta_2} x^2 \frac{1}{\theta_2 - \theta_1} dx = \frac{\theta_2^2 + \theta_1\theta_2 + \theta_1^2}{3}$$

Method of Moments

- We then find $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k$ to solve the following moment condition

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^j}_{\text{sample moments}} = \underbrace{m_j(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)}_{\text{population moments}}, \quad j = 1, 2, \ldots, k,$$

  - The solutions: $(\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$ are the MM estimators of $(\theta_1, \theta_2, \ldots, \theta_k)$
  - k unknown parameters with k moment conditions

Method of Moments

- The method of moments involves equating sample moments with population moments.
- We impose the condition so that the sample moment is equal to the population moment: the moment condition

Moment Conditions

$$m_j(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n} X_i^j \xrightarrow{p} E(X^j) = m^j(\theta)$$

靠近

where $\theta = (\theta_1, \theta_2, \ldots, \theta_k)$, $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_k)$

Method of Moments

- The basic idea behind this form of the method is to:
  (1) Equate the first sample moment $m_1 = \frac{1}{n}\sum_{i=1}^n X_i$ to the first theoretical moment $E(X)$.
  (2) Equate the second sample moment $m_2 = \frac{1}{n}\sum_{i=1}^n X_i^2$ to the second theoretical moment $E(X^2)$.
  (3) Continue equating sample moments m_j with the corresponding theoretical moments $E(X^j)$, $j = 3, 4, \ldots$ until you have as many equations as you have parameters.
  (4) Solve for the parameters.

Example

- Let
$$\{X_i\}_{i=1}^n \sim^{i.i.d.} U(\theta_1, \theta_2)$$
find the MMEs for θ_1 and θ_2
- Recall that the moments are
$$E(X) = m_1(\theta_1, \theta_2) = \frac{\theta_1 + \theta_2}{2}$$
$$E(X^2) = m_2(\theta_1, \theta_2) = \frac{\theta_2^2 + \theta_1\theta_2 + \theta_1^2}{3}$$

Example

- The moment conditions are

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i = \underbrace{E(X) = \frac{\hat{\theta}_1 + \hat{\theta}_2}{2}}_{m_1(\hat{\theta}_1,\hat{\theta}_2)}$$

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 = \underbrace{E(X^2) = \frac{\hat{\theta}_2^2 + \hat{\theta}_1\hat{\theta}_2 + \hat{\theta}_1^2}{3}}_{m_1(\hat{\theta}_1,\hat{\theta}_2)}$$

Example

- We can solve for $\hat{\theta}_1$ and $\hat{\theta}_2$ as

$$\hat{\theta}_1 = \bar{X} - \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

$$\hat{\theta}_2 = \bar{X} + \sqrt{\frac{3}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Method of Moments: Remarks

- Note that the Method of Moment Estimator is not unique.
  - Different moment conditions may obtain different estimators.
- In general, we use the first few moments for simplicity.

Method of Maximum Likelihood

- Assume

$$\{X_i\}_{i=1}^n \sim^{i.i.d.} f(x,\theta)$$

where $f(\cdot)$ is known but θ is an unknown parameter.

- Joint pmf/pdf (function of random sample)

$$f(x_1,\dots,x_n;\theta) = f(x_1;\theta)\cdots f(x_n;\theta) = \prod_i f(x_i;\theta)$$

- We can also call it a likelihood function of θ:

$$\mathcal{L}(\theta) = \prod_i f(x_i;\theta)$$

Maximum Likelihood Estimator (MLE)

- The maximum likelihood estimator $\hat{\theta}$

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta)$$

  - To find the value of θ such that the random sample $\{X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n\}$ is most likely to be observed.

Example

- There are 5 balls in the urn.
- Let μ denote the portion of blue balls in the urn, and $1-\mu$ be the portion of green balls in the urn.
  - μ is the unknown parameter
- The random sample is $\{X_1, X_2, \ldots, X_{10}\}$, where

$$X_i = \begin{cases} 1 & \text{If the ball is blue,} \\ 0 & \text{If the ball is green.} \end{cases}$$

Example

- It is clear that

$$X_i \sim^{i.i.d.} \text{Bernoulli}(\mu)$$

- Let

$$Y_{10} = X_1 + X_2 + \cdots + X_{10} = \sum_{i=1}^{10} X_i$$

Y_{10} represents the number of blue ball, and

$$Y_{10} \sim \text{Binomial}(10, \mu)$$

Likelihood Function

Consider the following two possible samples

- Sample 1: $Y_{10} = 7$

μ	$P(Y_{10} = 7) = \binom{10}{7}\mu^7(1-\mu)^3$
0	0
1/5	0.000786
2/5	0.042467
3/5	0.214991
4/5	0.201327
5/5	0

Likelihood Function

- Sample 2: $Y_{10} = 2$

μ	$P(Y_{10} = 2) = \binom{10}{2}\mu^2(1-\mu)^8$
0	0
1/5	0.301990
2/5	0.120932
3/5	0.010617
4/5	0.000074
5/5	0

Likelihood Function

	Sample 1: $Y_{10} = 7$	Sample 2: $Y_{10} = 2$
μ	$P(S_n^* = 7) = \binom{10}{7}\mu^7(1-\mu)^3$	$P(S_n^* = 2) = \binom{10}{2}\mu^2(1-\mu)^8$
0	0	0
1/5	0.000786	0.301990
2/5	0.042467	0.120932
3/5	0.214991	0.010617
4/5	0.201327	0.000074
5/5	0	0

Maximum Likelihood Estimator

- If $\mathcal{L}(\theta)$ is differentiable, then the MLE is the solution to:

$$\frac{\partial \mathcal{L}(\theta)}{\partial \theta} = 0$$

- Note that

$$\hat{\theta} = \arg\max \mathcal{L}(\theta) = \arg\max \log \mathcal{L}(\theta)$$

- So the MLE is also the solution to:

$$\frac{\partial \log \mathcal{L}(\theta)}{\partial \theta} = 0$$

where $\log \mathcal{L}(\theta)$ is called the log likelihood function

Examples

- Let $\{X_i\}_{i=1}^n \sim^{i.i.d.}$ Bernoulli(μ), then the likelihood function is

$$\mathcal{L}(\mu) = \prod_{i=1}^n \mu^{x_i}(1-\mu)^{1-x_i} = \mu^{\sum_i x_i}(1-\mu)^{n-\sum_i x_i}$$

- The log likelihood function is

$$\log \mathcal{L}(\mu) = \left(\sum_i x_i\right)\log \mu + \left(n - \sum_i x_i\right)\log(1-\mu)$$

  - It can be shown that the estimate is $\hat{\mu} = \frac{1}{n}\sum_i x_i$, and hence the estimator is

$$\hat{\mu} = \frac{1}{n}\sum_i X_i = \bar{X}$$

Important Property of MLEs: Invariance

Theorem

If $\hat{\theta}$ is the MLE of θ, and let $\tau(\theta)$ be a function of θ, then $\hat{\tau} = \tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.

- Example:
  - Given
$$\{X_i\}_{i=1}^{n} \sim^{i.i.d.} \text{Bernoulli}(\mu)$$
  - The MLE of $Var(X_1) = \mu(1-\mu)$ is
$$\widehat{Var(X_1)} = \hat{\mu}(1-\hat{\mu}) = \bar{X}(1-\bar{X})$$

Section 2

Evaluating Estimators

Criteria for Evaluating Estimators

- Unbiased
- Efficient
- Consistent

Unbiasedness

Definition (Unbiasedness)

$\hat{\theta}$ is unbiased if $E(\hat{\theta}) = \theta$

- Hence, bias can be defined as

$$B(\theta) = E(\hat{\theta}) - \theta$$

Unbiasedness

- Given $\{X_i\}_{i=1}^{n} \sim^{\text{i.i.d.}} (\mu, \sigma^2)$. By analogy principle,
  - $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$
  - $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$
- It can be shown that
  - $E(\bar{X}) = \mu$
  - $E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2$
- Hence, to obtain unbiased estimator for σ^2, let

$$S^2 = \left(\frac{n}{n-1}\right)\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n-1},$$

so that $E(S^2) = \sigma^2$.

Minimum Variance Unbiased Estimator (MVUE)

Definition (MVUE)

$\hat{\theta}$ is an MVUE of θ if and only if

- $E(\hat{\theta}) = \theta$
- $Var(\hat{\theta}) \leq Var(\hat{\theta}^*)$ for all $\hat{\theta}^*$ such that $E(\hat{\theta}^*) = \theta$

Efficient

Definition (Relatively Efficient Estimator)

Given two unbiased estimators: $\hat{\theta}$ and $\tilde{\theta}$. We say that $\hat{\theta}$ is more efficient than $\tilde{\theta}$ if

$$Var(\hat{\theta}) \leq Var(\tilde{\theta})$$

- $\{X_i\} \sim^{i.i.d.} (\mu, \sigma^2)$
- Consider the following two estimators

$$\bar{X} = \frac{1}{n}\sum_i X_i, \quad \tilde{X} = \frac{X_1 + X_{100}}{2}$$

  - Both $\bar{X}$ and $\tilde{X}$ are unbiased
  - But $Var(\bar{X}) < Var(\tilde{X})$ when $n > 2$.

Unbiased vs. Efficient

- Given two unbiased estimators, the one with smaller variance is better.
- How to compare
  - Unbiased estimators with higher variance vs. Biased estimators with lower variance

Mean Squared Error (MSE)

Definition (Mean Squared Error)

Mean Squared Error (MSE) is defined by

$$\text{MSE}(\hat{\theta}) \equiv E\left[(\hat{\theta} - \theta)^2\right]$$

- Note that

$$\text{MSE}(\hat{\theta}_n) = Var(\hat{\theta}_n) + (B(\theta))^2$$

- The estimator with smaller MSE is called an MSE efficient estimator.

Example

- Given $\{X_i\}_{i=1}^n \sim^{\text{i.i.d.}} N(\mu, \sigma^2)$
- Let

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}, \quad S^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1}, \quad \hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$$

- It can be shown that

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

- That is, compared to S^2, $\hat{\sigma}^2$ is MSE efficient.

Consitent

Definition (Consistent Estimator)

An estimator $\hat{\theta}$ is a consistent estimator of θ, if

$$\hat{\theta}_n \xrightarrow{p} \theta$$

- Example: $\{X_i\}_{i=1}^n \sim^{i.i.d.} (\mu, \sigma^2)$,
  - By WLLN
  $$\bar{X}_n \xrightarrow{p} \mu$$
  - By WLLN and CMT
  $$S_n^2 \xrightarrow{p} \sigma^2$$
  $$\hat{\sigma}_n^2 \xrightarrow{p} \sigma^2$$