Chapter 9 Parameter Estimation

► Goldberger, Ch. 11

Yale Note Ch. 13, 14

Wackerly et al. Ch. 8

1. Estimation

- Random sample $(y_1, y_2, \cdots, y_n)$ drawn from $f(y;\theta)$, where θ denotes parameter.

► Our interest: parameter θ.

(Example) $\theta = \mu,\ \sigma^2,\ \sigma_{XY},\ \alpha,\ \beta$.

► Our task: estimate θ, $\hat{\theta} = h(y_1, \cdots, y_n)$.

- What function of $h(\cdot)$ shall we take?

► $\mathbf{Y} = (Y_1, \cdots, Y_n)$: random vector.

► $\mathbf{y} = (y_1, \cdots, y_n)$: sample data.

► $T = h(\mathbf{Y})$: estimator.

► $t = h(\mathbf{y})$: estimate.

- All we know about θ: θ is an element of the parameter space Ω.

► Information: $\theta \in \Omega$ (Ω: parameter space).

► *<u>Estimation</u>*: an attempt to elicit information about θ.

► *<u>Estimator</u>*: a function of random variable ($T = h(\mathbf{Y}) \Rightarrow \hat{\theta} = h(Y_1, \cdots, Y_n)$).

► *<u>Estimate</u>*: a realization of $\hat{\theta}$ ($t = h(\mathbf{y})$).

- Chief question: how to choose the function $h(\cdot)$.

2. Analogy Principle

- Since a population parameter is a feature of population,

  to estimate parameter, use the corresponding features of the sample.

► Most natural rule for selecting an estimator.

① To estimate population moment, use the corresponding sample moment.

(Example)

For $\theta = \mu$, use $\hat{\theta} = \overline{Y}$.

For $\theta = \sigma^2$, use $\hat{\theta} = S^2$.

② To estimate a function of population moment, use a function of sample moment.

(Example 1)

For the population BLP slope, $\beta = \dfrac{\sigma_{XY}}{\sigma_X^2}$, use $\hat{\beta} = \dfrac{s_{XY}}{s_X^2}$.

For the population BLP intercept, $\alpha = \mu_Y - \beta\mu_Y$, use $\hat{\alpha} = \overline{Y} - \hat{\beta}\overline{X}$.

(Example 2) For multivariate BLP, $E\left(\mathbf{X}_2 \middle| \mathbf{X}_1\right) = \alpha + B'\mathbf{X}_1$ with $B = \Sigma_{11}^{-1}\Sigma_{12}$, $\alpha = \boldsymbol{\mu}_2 - B'\boldsymbol{\mu}_1$

Analogue estimation: $\hat{E}\left(\mathbf{X}_2 \middle| \mathbf{X}_1\right) = \hat{\alpha} + \hat{B}'\mathbf{X}_1$ with $\hat{B} = \hat{\Sigma}_{11}^{-1}\hat{\Sigma}_{12}$, $\hat{\alpha} = \hat{\boldsymbol{\mu}}_2 - \hat{B}'\hat{\boldsymbol{\mu}}_1$

③ To estimate $F(c) = P(Y \le c)$, use $\hat{F}(c) = \dfrac{\#\ of\ \ Y_i \le c}{\#\ of\ \ drawings}$.

④ To estimate population median, use the sample median.

⑤ θ=maximum, $\hat{\theta}$=sample maximum.

3. Criterion for an Estimator

- Let $T = h(Y_1, \cdots, Y_n)$ be a sample statistic with a pdf $g(t)$ and moments $E(T), V(T)$.
- Choosing an estimator T $\Rightarrow$ choosing a sampling distribution of T.
    - ► We would like $T = \theta$, but that ideal is unattainable.
    - ► Maybe, we ask T be close to θ.

(Definition) If $E(T) = \theta$, T is an ***unbiased*** estimator of θ.

► Even though $T \neq \theta$, in average $T = \theta$.

► $Bias \text{ of } T = E(T) - \theta$.

(Definition) An unbiased estimator $\hat{\theta}_1$ is ***more efficient*** than an unbiased estimator $\hat{\theta}_2$ if $V(\hat{\theta}_1) < V(\hat{\theta}_2)$.

(Example) $\theta = \mu$.

Estimator 1: X_1. Estimator 2: $\bar{X}$.

$E(X_1) = E(\bar{X}) = \mu$: both are unbiased.

$V(X_1) = \sigma^2$, $V(\bar{X}) = \dfrac{\sigma^2}{n}$.

So, $\bar{X}$ is more efficient than X_1.

► Efficient estimators are also refined to as *best unbiased estimators* or more precisely as *minimum variance unbiased estimators*.

(Definition) T is a *minimum variance unbiased estimator(MVUE)* of θ iff

① $E(T-\theta)=0$ for all θ.

② $V(T) \le V(T^*)$ for all T^* such that $E(T^*-\theta)=0$.

• *Uniformly Minimum Variance Unbiased(UMVU)* estimator.

► *Cramer-Rao Lower Bound(Cramer Rao Inequality)*

- In general, there might be a trade-off between bias and variance.

► A natural measure of distance between the random variable T and the parameter θ: mean squared error(MSE).

(Definition) The *mean squared error*(MSE) of an estimator is:

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

$$= V\left(\hat{\theta}\right) + \left\{Bias\left(\hat{\theta}\right)\right\}^2$$

► For unbiased estimator, $MSE = Variance$.

- Often, we are finding the minimum variance estimator <u>among linear function</u> of sample observations.

(Definition) An estimator $\hat{\theta}$ is a <u>*linear estimator*</u> if it is a linear function of sample observations, $\hat{\theta} = \sum_{i=1}^{n} a_i Y_i$.

(Definition) Let $\hat{\theta}$ be an estimator of θ of the form $\hat{\theta} = \sum_{i=1}^{n} a_i Y_i$ where a_i's are constant. If $E\left(\hat{\theta}\right) = \theta$, and $V\left(\hat{\theta}\right) \le V\left(\tilde{\theta}\right)$ when $\tilde{\theta}$ is any other linear and unbiased estimator, then $\hat{\theta}$ is the <u>*best linear unbiased* estimator</u>(BLUE) of θ or <u>*minimum variance linear unbiased estimator*</u>*(MVLUE)* of θ.

(Example 1) (Population mean) Let $Y_1, Y_2, \cdots, Y_n$ be a random sample from a population with $E(Y_i) = \mu$, $V(Y_i) = \sigma^2$. Find MVLUE of μ.

☞ (Sol) MVLUE(BLUE) of μ: $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}$.

(Exercise) Prove that $\bar{X}$ is MVLUE of μ.

(Theorem) In random sampling, sample size n, from any population, the sample mean is the minimum variance linear unbiased estimator(MVLUE) of the population mean.

(Example) (Population raw moment) Similarly, $\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i^s$ is the MVLUE of $\theta = E(Y^s)$.

(Definition) *Robustness*

An estimator is *robust* if its sampling distribution is not affected by violations of the underlying assumption.

(Examples of violations)

① Outlier in the data: Mean($\bar{X}$) vs. Median($\bar{M}$):

Note $V(\bar{X}) = \dfrac{\sigma^2}{n}$, $V\left(\bar{M}\right) = \dfrac{1}{n4\left\{f(0)\right\}^2}$.

So, if $\sigma^2 > \dfrac{1}{4\left\{f(0)\right\}^2}$, prefer median and if $\sigma^2 < \dfrac{1}{4\left\{f(0)\right\}^2}$, prefer mean.

② Misspecification of p.d.f.

③ Heterogeneity or dependence in the data.

4. Asymptotic Criteria

(Definition) T_n is a *consistent estimator* of θ if $T_n \xrightarrow{p} \theta (p\lim T_n = \theta)$.

(Definition) T_n is *asymptotically unbiased* if $\lim E(T_n) = \theta$.

(Definition) T_n is *asymptotically efficient* if other asymptotic variance exceeds the asymptotic variance of T_n.

(Definition) T_n is ***best asymptotically normal*(BAN)** estimator of θ iff

① $T_n \overset{a}{\sim} N\left(\theta, \dfrac{\phi^2}{n}\right)$, and

② $\phi^2 \le \phi^{*2}$ for all T_n^* such that $T_n^* \overset{a}{\sim} N\left(\theta, \dfrac{\phi^2}{n}\right)$.

- BAN criterion is the asymptotic version of the MVUE criterion.

5. Interval Estimation

• For a specific sample, the single value obtained for $\hat{\theta}$ is called a *point estimate*.

► So far, we have been concerned with point estimation of population parameter.

► One difficulty with point estimators: they do not convey a sense of the precision of the estimator.

► It can be useful to provide interval estimators based on the sampling distribution of the estimator.

► $\mu \in \overline{X} \pm \text{sampling error}$

► The estimate obtained will vary from sample to sample.

► There is some probability that it will be quite erroneous.

- The logic behind the interval estimate:

  ► Use the sample data to construct an interval.

  ► Expect this interval to contain the true parameter in some specified proportion of samples, or equivalently, *with some desired level of confidence*.

(1) Estimation of Means

$\boxed{\text{Case 1}}$: $(y_1, \cdots, y_n) \sim N(\mu, \sigma^2)$ with known variance σ^2.

Then the sampling distribution of the sample mean $\bar{Y}$ is,

$$Z = \frac{\bar{Y} - \mu}{\sigma / \sqrt{n}} \sim N(0,1).$$

Let $A = \{ |Z| \le 1.96 \}$,

so, $P(A) = \Phi(1.96) - \Phi(-1.96) = 0.975 - 0.025 = 0.95.$

$\Rightarrow$

$$A = \left\{ \left| \bar{Y} - \mu \right| \le 1.96\, \sigma / \sqrt{n} \right\}$$

$$= \left\{ \bar{Y} - 1.96\, \sigma / \sqrt{n} \le \mu \le \bar{Y} + 1.96\, \sigma / \sqrt{n} \right\}$$

► "The parameter μ lies in the interval $\bar{Y} \pm 1.96\, \sigma / \sqrt{n}$" is true with probability 95%.

► $\bar{Y} \pm 1.96\, \sigma / \sqrt{n}$ is 95% confidence interval for the parameter μ.

- $100(1-\alpha)\%$ confidence interval for μ: $\bar{Y} \pm z_{\alpha/2}\,\sigma/\sqrt{n}$.

☞

$\alpha = 0.05 \;\Rightarrow\; z_{\alpha/2}$

$\alpha = 0.10 \;\Rightarrow\; z_{\alpha/2}$

$\alpha = 0.01 \;\Rightarrow\; z_{\alpha/2}$

(Definition) The random interval $\left[\bar{Y} - z_{\alpha/2}\,\sigma/\sqrt{n}, \bar{Y} + z_{\alpha/2}\,\sigma/\sqrt{n}\right]$ is an *interval estimator* of μ.

(Definition) If we replace the random variable $\bar{Y}$ with an estimate based on the values of a particular sample, we obtain an *interval estimate*.

(Remark) A common mistake is to say that " the interval estimate contains the true value μ with probability $1-\alpha$." Note that once this particular interval estimate has been constructed, the true value μ must be either inside or outside the interval with certainty, so the statement cited does NOT make any sense.

► Rather the interval estimate is one realization of the interval estimator which we constructed to include μ in repeated trials with probability $1-\alpha$.

Case 2: $\left(y_1,\cdots,y_n\right) \sim N(\mu,\sigma^2)$ with unknown variance σ^2.

Then,

ⓐ $Z = \dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$.

ⓑ $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

ⓒ $\bar{Y}$ and S^2 are statistically independent.

Thus,

$$t = \frac{\dfrac{\bar{Y}-\mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{(n-1)S^2}{\sigma^2}\Big/(n-1)}} = \frac{\bar{Y}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

- The corresponding interval estimator is $\left[\bar{Y}-t_{(n-1:\alpha/2)}\, s/\sqrt{n}\,,\bar{Y}+t_{(n-1;\alpha/2)}\, s/\sqrt{n}\right]$.

- $s/\sqrt{n}$: standard error of $\bar{Y}$.
- $\sigma/\sqrt{n}$: standard deviation of $\bar{Y}$.

$\boxed{\text{Case 3}}$: $(y_1, \cdots, y_n)$: random sample from <u>unknown</u> distribution(unknown mean μ, unknown variance σ^2) with large sample.

- Since $Z = \dfrac{\bar{Y} - \mu}{s/\sqrt{n}} \xrightarrow{d} N(0,1)$,

the corresponding *approximate* $100(1-\alpha)\%$ *confidence interval* is

$\left[\bar{Y} - z_{\alpha/2}\, s/\sqrt{n}\,, \bar{Y} + z_{\alpha/2}\, s/\sqrt{n}\right]$.

► $s/\sqrt{n}$: asymptotic standard error of $\bar{Y}$.

(2) Estimation of Variances

$\boxed{\text{Case 1}}$: $\left(y_1,\cdots,y_n\right) \sim N(\mu,\sigma^2)$.

Since $\dfrac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$,

$$P\left[\chi^2_{(n-1;\alpha/2)} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{(n-1;1-\alpha/2)}\right] = 1-\alpha.$$

☞

So, the interval estimator is $\left[\dfrac{(n-1)s^2}{\chi^2_{(n-1;1-\alpha/2)}}, \dfrac{(n-1)s^2}{\chi^2_{(n-1;\alpha/2)}}\right]$.

$\boxed{\text{Case 2}}$: $(y_1,\cdots,y_n)\sim$unknown distribution.

- If the underlying distribution is not normal, the finite sample distribution of s^2 is unknown. However, we can use the asymptotic normal distribution of s^2.

So,

$$\frac{\sqrt{n}\left(s^2-\sigma^2\right)}{\left(\hat{\mu}_4-s^4\right)^{1/2}} \xrightarrow{d} \frac{\sqrt{n}\left(s^2-\sigma^2\right)}{\left(\hat{\mu}_4-\sigma^4\right)^{1/2}} \xrightarrow{d} N(0,1),$$

Where $\hat{\mu}_4=\dfrac{1}{n}\sum_{i=1}^{n}(Y_i-\overline{Y})^4$.

- Report $\left[s^2\pm 1.96\left(\hat{\mu}_4-s^4\right)^{1/2}\Big/\sqrt{n}\right]$ as the approximate 95% confidence interval for σ^2.