# CH 9. Properties of Point Estimators and Methods of Estimation

Jeongyoun Ahn

KAIST

# Relative Efficiency

- $\hat{\theta}_1$ and $\hat{\theta}_2$: unbiased estimators of a parameter $\theta$.
- The efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}$$

- If $V(\hat{\theta}_2) > V(\hat{\theta}_1)$, $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) > 1$: $\hat{\theta}_1$ is preferred to $\hat{\theta}_2$.

# Consistency

- $\hat{\theta}_n$ is said to be a consistent estimator of $\theta$, if, for any positive number $\epsilon$,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1$$

or, equivalently,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \epsilon) = 0.$$

- Notation:

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

Or, $\hat{\theta}_n$ converges to $\hat{\theta}$ in probability.

An unbiased estimator $\hat{\theta}_n$ for $\theta$ is a consistent estimator $\theta$ if

$$\lim_{n\to\infty} V(\hat{\theta}_n) = 0$$

# Consistency

If $\hat{\theta}_n \xrightarrow{p} \theta$ and $\hat{\theta}'_n \xrightarrow{p} \theta'$, then

- $\hat{\theta}_n + \hat{\theta}'_n \xrightarrow{p} \theta + \theta'$
- $\hat{\theta}_n \times \hat{\theta}'_n \xrightarrow{p} \theta \times \theta'$
- If $\theta' \neq 0, \hat{\theta}_n / \hat{\theta}'_n \xrightarrow{p} \theta / \theta'$
- If $g(\cdot)$ is a real-valued function that is continuous at $\theta$, then $g(\hat{\theta}_n) \xrightarrow{p} g(\theta)$

# Sufficiency

Motivation

- How do you come up with a form of an estimator $\hat{\theta}$?
- A **sufficient statistic** summarizes all the information in a sample about a target parameter $\theta$.
- A trivial sufficient statistic: collection of all $n$ observations $(Y_1, Y_2, \ldots, Y_n)$
- Can we find a "simpler" statistic that still contains all the information about $\theta$?
- A "Good" estimators should be a function of sufficient statistic.

# Sufficiency

Sufficiency

- $Y_i \overset{iid}{\sim} p(\cdot|\theta)$ (for discrete) or $f(\cdot|\theta)$ (for continuous) for $i = 1, \ldots, n$.
- Then, a statistic $U = g(Y_1, Y_2, \ldots, Y_n)$ is said to be **sufficient** for $\theta$ if the conditional distribution $Y_1, Y_2, \ldots, Y_n|U$ does not depend on $\theta$.
- e.g. $X_i \overset{iid}{\sim} B(1, p)$, then $Y = \sum_{i=1}^{n} X_i$ is sufficient for $p$.

# Sufficiency

Sufficient statistic (SS)

- Q: How to find a sufficient statistic? A: Using the "likelihood" function.
- $y_1, y_2, \ldots, y_n$: sample observations of random variables $Y_1, Y_2, \ldots, Y_n$.
- Likelihood is **joint** probability function (for discrete r.v.) and **joint** probability density function (for continuous r.v.)
  - Discrete r.v., $Y_i \overset{iid}{\sim} p(y|\theta)$: $L(y_1, y_2, \ldots, y_n|\theta) = p(y_1|\theta)p(y_2|\theta)\cdots p(y_n|\theta)$
  - Continuous r.v., $Y_i \overset{iid}{\sim} f(y|\theta)$: $L(y_1, y_2, \ldots, y_n|\theta) = f(y_1|\theta)f(y_2|\theta)\cdots f(y_n|\theta)$

# Factorization theorem

$U$: a statistic based on $Y_1, Y_2, \ldots, Y_n$.
Then, $U$ is a **sufficient statistic** for the estimation of $\theta$ if and only if the likelihood $L(\theta) = L(y_1, y_2, \ldots, y_n|\theta)$ can be factored into two functions,

$$L(y_1, y_2, \ldots, y_n|\theta) = g(u, \theta) \times h(y_1, y_2, \ldots, y_n)$$

where $g(u, \theta)$ is a function only of $u$ and $\theta$ and $h(y_1, y_2, \ldots, y_n)$ is not a function of $\theta$.

- Note that there can be many sufficient statistics.
- Minimal Sufficient Statistic (MSS): most "simplified" sufficient statistic.

# The Rao-Blackwell Theorem

- Sufficient statistics help us to find the minimum-variance estimator among unbiased estimators.
- (Theorem 9.5, p.464) $\hat{\theta}$: an unbiased estimator for $\theta$ and $V(\hat{\theta}) < \infty$.
  $U$: a sufficient statistic for $\theta$ and define $\hat{\theta}^* = E(\hat{\theta}|U)$.

  $$E(\hat{\theta}^*) = \theta \quad \text{and} \quad V(\hat{\theta}^*) \leq V(\hat{\theta})$$

- In words: An **unbiased** estimator ($\hat{\theta}$) can be improved (i.e., smaller variance) by becoming a function of a sufficient statistic ($U$).
- In turn, If $\hat{\theta}$ is a function of a (minimal) sufficient statistic to begin with, the Rao-Blackwell theorem implies that $\hat{\theta}$ is a Minimum Variance Unbiased Estimator (MVUE). (That is, it can not be further improved)

# The Method of Moments

The method of moments estimator (MME)

- Let $\mu'_k = E(Y^k)$ (population $k$th moment) and $m'_k = \frac{1}{n}\sum_{i=1}^n Y_i^k$ (sample $k$th moment).
- Idea: substitute sample moments, $m'_k$, into the moments of a probability distribution, $\mu'_k$.
- Why does MME work? Law of large numbers.

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^k \xrightarrow{p} E(Y^k) = \mu'_k$$

- Easy and intuitive, and consistent. However, not usually the best estimator. Can be biased.

# The Method of Maximum Likelihood

Maximum likelihood estimator (MLE)

- Urn example: sample two balls out of three balls (blue or red), $\theta =$ the number of red balls
  Assume that we observed two red balls, how to estimate $\theta$?
- $\hat{\theta}^{MLE}$: the value of $\theta$ that maximizes the likelihood $L(y_1, y_2, \ldots, y_n|\theta)$, or equivalently maximizes the log likelihood $\ell(y_1, y_2, \ldots, y_n|\theta) = \ln L(y_1, y_2, \ldots, y_n|\theta)$.
- Idea: makes the observed data "most probable" or "most likely"
- $\hat{\theta}$: MLE for $\theta$ and $t(\theta)$ is a function of $\theta$ that possesses a unique inverse. Then, $\widehat{t(\theta)}^{MLE} = t(\hat{\theta}^{MLE})$.

# Large-Sample Properties of MLE

- MLEs are consistent and asymptotically unbiased under some conditions on the probability function.
- The sampling distribution of MLE is approximately normal under some conditions on the probability function, i.e.

$$\sqrt{nI(\theta_0)}(\hat{\theta}^{MLE} - \theta_0) \xrightarrow{d} N(0,1)$$

or, equivalently,

$$\hat{\theta}^{MLE} \approx N\left(\theta_0, \frac{1}{nI(\theta_0)}\right)$$

where the Information function $I(\theta)$ is defined by

$$I(\theta) = E\left[\frac{\partial}{\partial\theta} \ln f(Y|\theta)\right]^2 = -E\left[\frac{\partial^2}{\partial\theta^2} \ln f(Y|\theta)\right]$$