# Goodness of Estimators

# Learning Outcomes

- Consistency
- Sufficiency
- Information
- Efficiency

# Consistency

An estimator is considered a consistent estimator of $\theta$ if the estimator, on average, converges to $\theta$ as $n \to \infty$.

# Consistency

Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter $\theta$. The estimator $\hat{\theta}$ is a consistent estimator of the $\theta$ if

1. $E\{(\hat{\theta} - \theta)^2\} \rightarrow 0$ as $n \rightarrow \infty$
2. $P(|\hat{\theta} - \theta| \geq \epsilon) \rightarrow 0$ as $n \rightarrow \infty$ for every $\epsilon > 0$

# Sufficiency

Sufficiency evaluates whether a statistic (or estimator) contains enough information of a parameter $\theta$. In essence a statistic is considered sufficient to infer $\theta$ if it provides enough information about $\theta$.

# Sufficiency

Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameter $\theta$. A statistic $T = t(X_1, \ldots, X_n)$ is said to be sufficient for making inferences of a parameter $\theta$ if condition joint distribution of $X_1, \ldots, X_n$ given $T = t$ does not depend on $\theta$.

# Factorization Theorem

The **Factorization Theorem** provides a condition for a statistic $T(X)$ to be sufficient for a parameter $\theta$ given a probability density function or probability mass function.

# Factorization Theorem

Let $X = (X_1, X_2, \dots, X_n)$ be a random sample with joint probability density (or mass) function $f(x|\theta)$, where $\theta$ is a parameter.

**Theorem**: A statistic $T(X)$ is sufficient for $\theta$ if and only if the joint density (or mass) function $f(x|\theta)$ can be factored into the form

$$f(x|\theta) = g(T(x), \theta) \cdot h(x)$$

where:

- $g(T(x), \theta)$ is a function that depends on $T(x)$

# Factorization Theorem

In other words, $f(x|\theta)$ can be written as a product of two functions, where only one function depends on the parameter $\theta$ and the sufficient statistic $T(X)$.

Implications: The Factorization Theorem is useful for identifying sufficient statistics, which summarize all necessary information from a sample about the parameter $\theta$.

# Example

Let $X_1, \ldots, X_n \overset{iid}{\sim} Bernoulli(p)$ and $Y_n = \sum_{i=1}^{n} X_i$. Show that $Y_n$ is a sufficient statistic for $p$.

# Example

Let $X_1, \dots, X_n \overset{iid}{\sim} Normal(\mu, \sigma^2)$ and $Y_n = \sum_{i=1}^{n} X_i$. Show that $Y_n$ is a sufficient statistic for $\mu$. Assume $\sigma^2$ is known.

# Information

In Statistics, information is thought of as how much does the data tell you about a parameter $\theta$. In general, the more data is provided, the more information is provided to estimate $\theta$.

# Information

Information can be quantified using Fisher's Information $I(\theta)$. For a single observation, Fisher's Information is defined as

$$I(\theta) = E\left[-\frac{\partial^2}{\partial\theta^2}\log\{f(X;\theta)\}\right],$$

where $f(X;\theta)$ is either the PMF or PDF of the random variable $X$.

# Information

Furthermore, $I(\theta)$ can be defined as

$$I(\theta) = Var\left\{\frac{\partial}{\partial\theta}\log f(X;\theta)\right\}.$$

# Proof

Show the following property:

$$E\left[-\frac{\partial^2}{\partial\theta^2}\log\{f(X;\theta)\}\right] = Var\left\{\frac{\partial}{\partial\theta}\log f(X;\theta)\right\}$$

# Efficiency

Efficiency of an estimator $T$ is the ratio of variation compared to the lowest possible variance.

# Efficiency

The efficiency of an estimator $T$, where $T$ is an unbiased estimator of $\theta$, is defined as

$$efficiency\, of\, T = \frac{1}{Var(T)nI(\theta)}$$

# Example

Let $X_1, \dots, X_n \overset{iid}{\sim} Unif(0, \theta)$ and $\widehat{\theta} = 2\bar{X}$. Find the efficiency of $\widehat{\theta}$.