2022 SUMMER

업데이터 통계학 스터디

Chapter 7 – Point Estimation 2

Sufficiency

Sufficiency

Ex. $Y_1, \cdots, Y_n$: $iid\ Bernoulli(p)$

Inference about p uses count of successes $S = \sum_1^n Y_i \sim Bin(n, p)$

Do we lose information about p in going from n observations to 1 sum?

Sufficiency

Ex. $Y_1, \cdots, Y_n$: $iid\ Unif(0, \theta)$, Good estimator based on W $= Y_{(n)}$

Does W contain all information about θ available from data?

Data reduction: $Y_1, \cdots, Y_n \rightarrow S = \sum_1^n Y_i$ or W $= Y_{(n)}$

Idea: A sufficient statistic compresses data without losing information about the parameter.

Sufficient Statistics

Definition (SS)

$Y_1, \cdots, Y_n$: random sample from a distribution with unknown θ.

A statistic $U = U(Y_1, \cdots, Y_n)$ is a sufficient statistic if conditional distribution of $Y_1, \cdots, Y_n$ given U does not depend on θ.

Meaning: if you already know $U = U(Y_1, \cdots, Y_n)$, any other statistic does not have any extra information about θ.

Example 8.1

$Y_1, \cdots, Y_n$: $iid\ Bernoulli(p)$

Let $S = \sum_1^n Y_i \sim Bin(n, p)$. Find a distribution of $Y_1, \cdots, Y_n | S$?

Likelihood Function

Definition

$y_1, \cdots, y_n$: sample observations taken on corresponding random variables $Y_1, \cdots, Y_n$ whose distributions depend on θ. The likelihood function of the sample is

$$L(\theta | y_1, \cdots, y_n) = f(y_1 | \theta) \times \cdots \times f(y_n | \theta)$$

Likelihood = joint probability/density function of $Y_1, \cdots, Y_n$

Example 8.2

$Y_1, \cdots, Y_n$: $iid\ Exponential(\theta)$

$L(\theta | y_1, \cdots, y_n) = ?$

Factorization Theorem

Theorem

$U = U(Y_1, \cdots, Y_n)$ is a SS for θ iff we can write the likelihood in the form as

$$L(\theta) = g(u(y_1, \cdots, y_n), \theta)h(y_1, \cdots, y_n)$$

Example 8.3

$Y_1, \cdots, Y_n$: $iid\ Exponential(\theta)$

Find a SS for θ

Example 8.4

$Y_1, \cdots, Y_n$: $iid\ Geometric(p)$

Find a SS for p

Example 8.5

$Y_1, \cdots, Y_n$: $iid\ Bernoulli(p)$

Find a SS for p

Example 8.6

$Y_1, \cdots, Y_n$: $iid\ Uniform(0, \theta)$

Find a SS for θ

Example 8.7

$Y_1, \cdots, Y_n$: $iid\ Gamma(\alpha, \beta)$

1) α is known, Find a SS for β
2) β is known, Find a SS for α
3) α and β are both unknown, SS for (α, β)

Example 8.8

$Y_1, \cdots, Y_n$: $iid\ N(\mu, \sigma^2)$

1) σ^2 is known, Find a SS for μ
2) μ is known, Find a SS for σ^2
3) μ and σ^2 are both unknown, SS for (μ, σ^2)

Sufficient Statistics

Notes:

- Any 1-1 function of a sufficient statistic is a sufficient statistic.

  Ex. $Exp(\theta): U = \sum Y_i, V = \frac{1}{n}\sum Y_i$

- Any statistic from which a sufficient statistic is calculated is also a sufficient statistic. Ex) Random sample itself

- Many possible SS's $\Rightarrow$ MSS(Minimal Sufficient Statistics)