Optimal estimation theory: Sufficiency

Dr. Qiuzhuang Sun

STAT3023

Setup (Review from STAT2011/2911)

- $X_1, \ldots, X_n$ is a random sample from a population distribution $f_X(\theta)$, i.e., $X_i \sim f_X(x;\theta)$ and they are mutually independent.
- The parameter θ characterizes the distribution, but unknown.
- Estimation problem: construct a function of $X_1, \ldots, X_n$ that provides information about θ.
    - ◇ Any function of $X_1, \ldots, X_n$ is called a **statistic**.
    - ◇ The statistic that is used to provide information about θ is called an **estimator** of θ, typically denoted as $\hat{\theta}$. (Rule to compute from data.)
    - ◇ The realized value of an estimator is called an estimate. (Applying the rule to specific data.)
- Example: The sample mean $\bar{X}$ is an estimator. After seeing the data $(X_1, X_2, X_3) = (5, 3, 4)$, the estimate is $\frac{5+3+4}{3} = 4$.

Methods of finding estimators (Review)

- Method-of-moment: Equate the sample moments $m_k = n^{-1} \sum_{i=1}^{n} X^k$ with the population moments $E(X^k) = g(\theta)$, and solve for θ.

- Maximum likelihood: Find θ that maximizes the likelihood function

$$L(\theta) = \prod_{i=1}^{n} f_X(X_i; \theta).$$

This is typically done via maximizing the log-likelihood

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_X(X_i; \theta).$$

Example

Estimate the parameters of $N(\mu, \theta)$ using the method-of-moment and maximum likelihood estimation.

Example

Example

Sufficiency

Sufficiency

A statistic $T(X) = T(X_1, \dots, X_n)$ is a sufficient statistic for θ if the conditional distribution of X given the value of $T(X)$ does not depend on θ.

Example: Let $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Show $T(X) = \sum_{i=1}^n X_i$ is a sufficient statistic of p.

Sufficiency

(Example continued)

Sufficiency

Example: Assume that the probability of seeing a head when flipping a coin is p. We toss the coin n times. Let $X_i = 1$ if the ith toss is head and $X_i = 0$ otherwise, $i = 1, \ldots, n$. Use $X_1, \ldots, X_{10}$ to estimate p.

We showed $T(X) = \sum_{i=1}^{n} X_i$ is a sufficient statistic of p for $X_1, \ldots, X_n \sim \text{Bernoulli}(p)$.

Sufficiency

Example: Let $X_1, \ldots, X_n$ be independent $N(\theta, 1)$. Show $\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Sufficiency

Example: Let $X_1, \ldots, X_n$ be independent $N(\theta, 1)$. Show $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a sufficient statistic for θ.

Sufficiency

Hence, any information about θ from the data $X_1, \ldots, X_n$ has to go through $T(X)$.

Sufficiency and likelihood function

Neyman factorization theorem: $T(X)$ is sufficient statistic for θ if and only if the likelihood function is written in the following form:

$$L(\theta; X) = g(T(X); \theta) h(X).$$

Proof:

Sufficiency and likelihood function

(Proof continued)

Sufficiency and likelihood function

(Proof continued)

Sufficiency and likelihood function

Example: Use the Neyman factorization theorem to find a sufficient statistic for $X_1, \ldots, X_n \sim \text{Bern}(p)$.

Sufficiency and likelihood function

Example: Use the Neyman factorization theorem to find a sufficient statistic for $X_1, \ldots, X_n \sim N(\theta, 1)$.

Sufficiency and likelihood function

Example: Use the Neyman factorization theorem to find a sufficient statistic for $X_1, \ldots, X_n \sim \text{Poisson}(\lambda)$.

Sufficient statistics in an exponential family

Suppose $X_1, \ldots, X_n$ is a random sample from an exponential family distribution with the pdf or pmf in the form

$$f(x|\theta) = h(x) \exp\left(\sum_{i=1}^{k} w_i(\theta) t_i(x) - A(\theta) \right).$$

Then a sufficient statistic for θ is

$$T(X) = [t_1(X), \ldots, t_k(X)] = \left[\sum_{i=1}^{n} t_1(X_i), \ldots, \sum_{i=1}^{n} t_1(X_i) \right].$$

Sufficient statistics in an exponential family

Example: Revisit the sufficient statistic for $X_1, \ldots, X_n \sim N(\theta, 1)$.

In general, the sufficient statistic is NOT unique, but it is unique up to scale changes.