Optimal estimation theory

Dr. Linh Nghiem

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 is 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 $\hat{\theta}$.
  - ◇ The realized value of an estimator is called an estimate.

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

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

Sufficiency

Sufficiency

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

Eg: $X_1, \ldots, X_n \sim$ Bernoulli(p) and $T(X) = \sum_{i=1}^{n} X_i$.

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:

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

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

Then a sufficient statistic for θ is

Consistency, unbiasedness, and efficiency

Consistency

An estimator $\hat{\theta}$ is consistent for θ if $\hat{\theta} \xrightarrow{p} \theta$ when $n \to \infty$; in other words, for any $\varepsilon > 0$, we have

$$\lim_{n \to \infty} P\left(|\hat{\theta} - \theta| \geq \varepsilon\right) = 0.$$

Consistency and mean square error (MSE)

Bias–variance decomposition of mean square error

Unbiased estimator

An estimator $\hat{\theta}$ is unbiased for θ if and only if

$$\text{Bias}(\hat{\theta}) = E(\hat{\theta}) - \theta = 0.$$

Hence, an unbiased estimator is consistent if

$$\mathsf{Var}(\hat{\theta}) \to 0, \text{ as } n \to \infty.$$

Efficiency

Given the two unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$, the relative efficiency of $\hat{\theta}_1$ versus $\hat{\theta}_2$ is defined as

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

Cramer-Rao Lower Bound (CRLB)

Attainment of CRLB

Sufficiency and unbiasedness

(Rao-Blackwell Theorem) Let $\hat{\theta}$ be an unbiased estimator for θ, and let T be a sufficient statistic for θ. Define $\hat{\theta}_2 = E(\hat{\theta} \mid T)$. Then $\hat{\theta}_2$ is unbiased for θ and is a uniformly more efficient than $\hat{\theta}_1$.