Goodness of Estimators

Learning Outcomes

- Cramér-Rao Inequality
- Relative Efficiency
- Consistency
- Sufficiency

Efficiency

The efficiency of an estimator T, where T is an unbiased estimator of θ, is defined as

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

Relative Efficiency

Given 2 unbiased estimators $\hat{\theta}_1$ and $\hat{\theta}_2$ of a parameter θ , with variances $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$, respectively, then the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is defined as

$$releff(\hat{\theta}_1, \hat{\theta}_2) = \frac{\hat{\theta}_1}{\hat{\theta}_2}$$

Consistency

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

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

Weak Law of Large Numbers

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, for every, $\epsilon > 0$,

$$\lim_{n\to\infty} P(|\bar{X} - \mu| < \epsilon) = 1$$

that is, $\bar{X}_n$ converges in probability to μ.

Strong Law of Large Numbers

Let $X_1, \ldots, X_n$ be iid random variables with $E(X_i) = \mu$ and $Var(X_i) = \sigma^2 < \infty$. Let $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$, for every, $\epsilon > 0$,

$$P(\lim_{n \to \infty} |\bar{X} - \mu| < \epsilon) = 1$$

that is, $\bar{X}_n$ converges almost surely to μ.

Cramér-Rao Inequality

Let $f(x_1, \dots, x_n; \theta)$ be the joint PMF of PDF of $X_1, \dots, X_n$ and $T = t(X_1, \dots, X_n)$ be an unbiased estimator of θ. Then

$$Var(T) \geq \frac{1}{nI(\theta)}$$

If $Var(T) = \frac{1}{nI(\theta)}$, then T is considered the most efficient estimator of θ.

Sufficiency

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

Sufficiency

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

Joint Sufficient Statistics

Let $X_1, \ldots, X_n$ be a random sample from a distribution with parameters $\theta = (\theta_1, \ldots, \theta_m)^{\mathrm{T}}$. A joint statistic $T = \{t_1(X_1, \ldots, X_n), \ldots, t_k(X_1, \ldots, X_n)\}^{\mathrm{T}}$ is said to be sufficient for making inferences on paramters θ if condition joint distribution of $X_1, \ldots, X_n$ given $T = t$ does not depend on θ.

Factorization Theorem

Let $f(x_1, \dots, x_n; \theta)$ be the joint PMF of PDF of $X_1, \dots, X_n$. Then $T = t(X_1, \dots, X_n)$ is a sufficient statistic for θ if and only if there exist 2 nonnegative functions, g and h, such that

$$f(x_1, \dots, x_n) = g\{t(x_1, \dots, x_n); \theta\}h(x_1, \dots, x_n).$$

Minimum Sufficient Statistics

A minimum sufficient statistic is a sufficient statistic that has the smallest dimensionality, which represents the greatest possible reduction of the data without any information loss.

Complete Sufficient Statistics

Let $f(t;\theta)$ be the joint PMF of PDF of $T = t(X_1, \dots, X_n)$. The family distribution is called complete if $E_\theta\{g(T)\} = 0$ for all θ implies $P_\theta\{g(T)\} = 1$ for all θ. Then T is called a complete sufficient statistic.

Example 1

Let $X_1, \dots, X_n \overset{iid}{\sim} Pois(\lambda)$, show that $\sum_{i=1}^{n} X_i$ is a sufficient statistic for λ.

Example 2

Let $X_1, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$, show that $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)^{\mathrm{T}}$ is a sufficient statistic for μ and σ^2.