AID-521 Mathematics for Data Science

Module: Statistics | Lecture: 4

POINT ESTIMATION & PROPERTIES

bias, mean square error, consistency, efficiency, sufficiency

Estimation Methods

Standard Methods Available

- → **Method of moments**
- → **Method of maximum likelihood**
- → Bayes' method
- → Generalized method of moments (semi-parametric)

Properties of Estimators

Criteria for choosing among various methods

Depends on objective of statistical analysis

→ **Unbiasedness**

→ **Consistency**

→ **Efficiency**

→ Sufficiency

Unbiased Estimators

A point estimator $\hat{\theta}$ is called an unbiased estimator of population parameter θ if

$$E[\hat{\theta}] = \theta$$

The bias of $\hat{\theta}$ is

$$E[\hat{\theta}] \; - \; \theta \quad := \; B(\hat{\theta})$$

Unbiased Estimators

- → mean

$$E[\bar{X}] = E[\frac{1}{n}\sum_{i=1}^{n} X_i] = \mu$$

- → variance

$$E[S^2] = E[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2] = \sigma^2$$

- → convex combinations

$$\ldots$$

Mean Square Error

The mean square error of estimator $\hat{\theta}$ is

$$MSE(\hat{\theta}) = E[\, (\hat{\theta} - \theta)^2 \,]$$

$\rightarrow$ $MSE(\hat{\theta}) = Var(\hat{\theta}) + (B(\hat{\theta}))^2$

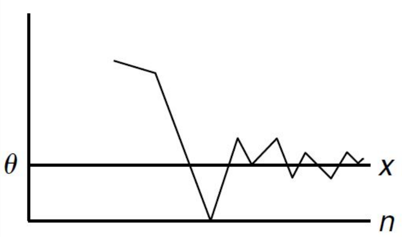
$\rightarrow$ MVUE† of θ is the unbiased $\hat{\theta}$ that minimizes MSE

† minimum variance unbiased estimator

Consistent Estimators

$\hat{\theta}_n$ is a consistent estimator of θ if

$$\lim_{n \to \infty} P(|\hat{\theta}_n - \theta| \leq \epsilon) = 1 \quad \forall \epsilon > 0$$



Consistent Estimators

Let $\hat{\theta}_n$ be an estimator of θ and let $Var(\hat{\theta}_n) < \infty$.

If

$$\lim_{n \to \infty} E[\,(\hat{\theta}_n \;-\; \theta)^2\,] \;=\; 0$$

then $\hat{\theta}_n$ is a consistent estimator of θ.

Consistent Estimators

Let $X_1, ..., X_n$ be a random sample from $\mathcal{N}(\mu, \sigma^2)$ population.

→ Show that the sample variance S^2 is a consistent estimator for σ^2.

→ Show that the maximum likelihood estimators for μ and σ^2 are consistent estimators for μ and σ^2.

Efficient Estimators

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are two <u>unbiased</u> estimators for θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is the ratio

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

If $e(\hat{\theta}_1, \hat{\theta}_2) > 1$, then $\hat{\theta}_1$ is relatively more efficient than $\hat{\theta}_2$.

Efficient Estimators

An estimator $\hat{\theta}_1$ is more efficient than $\hat{\theta}_2$ if

$$MSE(\hat{\theta}_1) \;\leq\; MSE(\hat{\theta}_2)$$

with strict inequality for some θ.

Efficient Estimators

UMVUE – uniformly minimum variance unbiased estimator

An unbiased estimator $\hat{\theta}_0$ is a UMVUE for θ if, for any other unbiased estimator $\hat{\theta}$

$$Var(\hat{\theta}_0) \;\leq\; Var(\hat{\theta})$$

for all possible values of θ.

Lower bound made available by Cramer-Rao inequality

Sufficient Estimators

$X_1, \ldots, X_n$ is a random sample from a distribution with unknown parameter θ.

Statistic $U = g(X_1, \ldots, X_n)$ is said to be sufficient for θ if the conditional pdf of $X_1, \ldots, X_n$ given $U = u$ does not depend on θ for any value of u.

An estimator of θ that is a function of a sufficient statistic for θ is said to be a sufficient estimator of θ.