Sufficiency, Rao–Blackwell, and Lehmann–Scheffé

Statistics · Module 3 — Point Estimation

Node 3.5

Position in Knowledge Graph

System:
Statistics

Structural Domain:
Point Estimation

Node:
3.5

Core Definitions

Statistic

A **statistic** is any measurable function of the data:

$$T = T(X_1, \dots, X_n).$$

Sufficiency (Factorization Theorem)

A statistic T is **sufficient** for θ if the likelihood can be factored as:

$$f_\theta(x_1, \dots, x_n) = g_\theta(T(x_1, \dots, x_n))\; h(x_1, \dots, x_n),$$

where h does not depend on θ.

Equivalently: conditional distribution of the data given T does not depend on θ.

Intuition

A sufficient statistic is a *lossless compression of the data* for the purpose of estimating θ.

Once you know T, the remaining details of the sample contain no additional information about θ.

This is why exponential families matter: they often admit low-dimensional sufficient statistics.

Structural Role

This node gives you the canonical pipeline:

1. Find a sufficient statistic T.
2. Improve any unbiased estimator by conditioning on T (Rao–Blackwell).
3. If T is complete, the improved estimator is uniquely optimal (Lehmann–Scheffé).

This is "optimality by structure," not by brute-force search.

Theorem 1: Rao–Blackwell

Let $\hat{\theta}$ be an estimator of $g(\theta)$ with finite variance, and let T be sufficient for θ.

Define the Rao–Blackwellized estimator:

$$\hat{\theta}^{\star} = \mathbb{E}_{\theta}[\hat{\theta} \mid T].$$

Then:

1. If $\hat{\theta}$ is unbiased for $g(\theta)$, so is $\hat{\theta}^{\star}$.
2. Variance improves (or stays equal):

$$\mathrm{Var}_{\theta}(\hat{\theta}^{\star}) \leq \mathrm{Var}_{\theta}(\hat{\theta}).$$

Theorem 2: Completeness and Lehmann–Scheffé

A statistic T is **complete** if:

$$\mathbb{E}_{\theta}[a(T)] = 0 \text{ for all } \theta \quad \Rightarrow \quad \mathbb{P}(a(T) = 0) = 1.$$

Lehmann–Scheffé Theorem

If:

- T is sufficient and complete for θ,
- $\hat{\theta}$ is any unbiased estimator of $g(\theta)$,

then:

$$\hat{\theta}^{\star} = \mathbb{E}[\hat{\theta} \mid T]$$

is the **unique** uniformly minimum-variance unbiased estimator (UMVU) of $g(\theta)$.

Worked Example: Bernoulli(p)

If $X_i \sim \text{Bernoulli}(p)$,

the likelihood is:

$$L(p) = p^{\sum x_i}(1-p)^{n-\sum x_i}.$$

Thus:

$$T = \sum_{i=1}^{n} X_i$$

is sufficient for p.

An unbiased estimator of p is $\bar{X}$, which is already a function of T, so Rao–Blackwell does not change it.

This illustrates a key point:

> Once you are already using a sufficient statistic, there is no "information left" to gain.

Cross-Links

- **AI:** Sufficient representations = no loss of task-relevant information (representation learning analogue).
- **Linear Algebra:** Conditioning is projection in L^2 (variance reduction via orthogonal projection).
- **Economics:** Sufficient statistics appear in exponential family and discrete choice models.
- **Quantum:** Informationally complete measurements vs sufficient summaries for parameter estimation.

Structural Compression

Invariant:
Conditioning on sufficient information cannot increase variance.

Mechanism:
Rao–Blackwell is variance reduction via conditional expectation (projection).

Cross-System Echo:
In ML, better representations reduce sample complexity and variance.

Exercises

1. Prove the variance reduction statement using the law of total variance.
2. Show that if $\hat{\theta}$ is a function of T, then Rao–Blackwell leaves it unchanged.
3. For $X_i \sim \text{Poisson}(\lambda)$, show that $T = \sum X_i$ is sufficient for λ.
4. Explain (conceptually) why completeness yields uniqueness in Lehmann–Scheffé.

Statistics · Module 3 — Point Estimation · Node 3.5
Concepts: sufficient-statistics, likelihood, conditional-probability
Prerequisites: 2.6, 3.3
Enables: 3.6

hbar.university — own.your.cognition