

Statistics for Economists II

Week 5: The Rao–Blackwell Theorem & MVUE

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Review: The Estimation Toolbox So Far
- 2 The Cramér–Rao Lower Bound
- 3 Sufficient Statistics and the Rao–Blackwell Theorem
- 4 Minimum Variance Unbiased Estimation (MVUE)
- 5 Completeness and the Lehmann–Scheffé Theorem
- 6 Worked Examples: Finding the MVUE
- 7 Why Does Conditioning on Sufficiency Help?
- 8 Connecting It All: The MVUE Roadmap
- 9 The Cramér–Rao Connection
- 10 Practice Problems
- 11 Summary and Big Picture

Where We Are: From Data to the Best Estimator

Over the past four weeks we have built a powerful toolkit:

- ① **Week 2:** Point estimation—introducing estimators

Economics Parallel

Finding the best estimator is like finding the most **efficient market**: we want to extract *every bit* of useful information from the data, with no waste.

Where We Are: From Data to the Best Estimator

Over the past four weeks we have built a powerful toolkit:

- ① **Week 2:** Point estimation—introducing estimators
- ② **Week 3:** Properties—unbiasedness, efficiency, MSE

Economics Parallel

Finding the best estimator is like finding the most **efficient market**: we want to extract *every bit* of useful information from the data, with no waste.

Where We Are: From Data to the Best Estimator

Over the past four weeks we have built a powerful toolkit:

- ① **Week 2:** Point estimation—introducing estimators
- ② **Week 3:** Properties—unbiasedness, efficiency, MSE
- ③ **Week 4:** Sufficiency—statistics that capture *all* information about θ

Economics Parallel

Finding the best estimator is like finding the most **efficient market**: we want to extract *every bit* of useful information from the data, with no waste.

Where We Are: From Data to the Best Estimator

Over the past four weeks we have built a powerful toolkit:

- ① **Week 2:** Point estimation—introducing estimators
- ② **Week 3:** Properties—unbiasedness, efficiency, MSE
- ③ **Week 4:** Sufficiency—statistics that capture *all* information about θ
- ④ **Today:** We combine these ideas to find the **best possible** unbiased estimator

Economics Parallel

Finding the best estimator is like finding the most **efficient market**: we want to extract *every bit* of useful information from the data, with no waste.

Quick Recall: The Three Pillars

Unbiasedness

$$\mathbb{E}(\hat{\theta}) = \theta \text{ for all } \theta$$

The estimator is *right on average*—no systematic error.

Sufficiency

T is sufficient for θ if the conditional distribution $f(X | T)$ does not depend on θ

T captures *all* information about θ .

Efficiency

Among unbiased estimators, the one with the **smallest variance** is most efficient

Less variability = more precise.

Today's big question: Can we **systematically** combine an unbiased estimator with a sufficient statistic to produce the **best** unbiased estimator? The answer is *yes*—and the recipe is the Rao–Blackwell Theorem.

Estimators Are Random Variables: The Height Example

We want to estimate μ = the average height in a population using a random sample.

The sample mean $\bar{X}$ is an estimator of μ :

- It is a **random variable**—it changes with every new sample
- Three people with heights 59, 63, 64 in:
 $\bar{x} = 62$
- Three different people: $\bar{x} = 64$ — the estimate changed!

$\bar{X}$ is **unbiased**: its distribution is always centered at the true μ , no matter which population we are in.

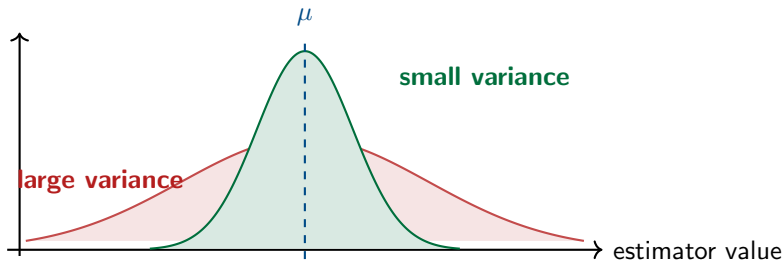
Being unbiased is good—but it is not the whole story. We also need to ask: **how spread out** are the estimates around μ ?

Unbiasedness: True $\mu = 65$

Region	Average height
Shortland	60 in
Mediumland	65 in
Tallland	70 in

Low Variance Is Better—But How Low Can It Go?

Two estimators, **both unbiased** (both centered at μ):



The **green** estimator is better: its estimates stay **closer to the true μ** .

Could we reach **zero variance**? Only by ignoring the data entirely (e.g. always guessing 65)—but that estimator is biased in Shortland and Tallland. **Not acceptable.**

The Key Question

Among all **unbiased** estimators, is there a **lowest possible variance**?

Yes—and that floor is the Cramér–Rao Lower Bound.

What Is Fisher Information?

The CRLB depends on the **Fisher information** $I(\theta)$: a number that tells us how much information *one observation* carries about θ .

Definition

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2} \log f(X; \theta)\right]$$

Peaked distribution $\Rightarrow$ high $I(\theta)$

The log-likelihood has a **sharp peak**: the data point strongly to one value of θ . Each observation is very informative.

Flat distribution $\Rightarrow$ low $I(\theta)$

The log-likelihood is **flat**: many values of θ fit the data equally well. Each observation tells us little.

Key scaling rule

For n i.i.d. observations, total Fisher information is $I_n(\theta) = n \cdot I(\theta)$. More data = more information.

Computing $l(\mu)$ for the Normal — Steps 1 & 2

Setting: $X \sim N(\mu, \sigma^2)$, σ^2 known.

Step 1 — Write the likelihood for one observation:

$$f(x; \mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right)$$

Step 2 — Take the log and drop anything not involving μ :

$$\ell(\mu) = -\frac{(x - \mu)^2}{2\sigma^2} + \text{const}$$

Taking the **log** turns the product of densities into a sum, making derivatives tractable. Dropping constants (like $-\frac{1}{2} \log 2\pi\sigma^2$) is safe because they vanish when we differentiate.

Computing $I(\mu)$ for the Normal — Steps 3, 4 & 5

Step 3 — First derivative (the *score function*):

$$\frac{\partial \ell}{\partial \mu} = \frac{x - \mu}{\sigma^2}$$

Step 4 — Second derivative:

$$\frac{\partial^2 \ell}{\partial \mu^2} = -\frac{1}{\sigma^2}$$

Step 5 — Fisher information (take the negative expectation):

$$I(\mu) = -\mathbb{E}\left[\frac{\partial^2 \ell}{\partial \mu^2}\right] = -\mathbb{E}\left[-\frac{1}{\sigma^2}\right] = \frac{1}{\sigma^2}$$

Result for n i.i.d. observations

$$I_n(\mu) = n \cdot \frac{1}{\sigma^2} = \frac{n}{\sigma^2}$$

In Step 5, the term $-1/\sigma^2$ contains **no** X , so its expectation is just itself. The algebra is simple—the setup is the hard part.

The Cramér–Rao Lower Bound

Theorem (Cramér–Rao Lower Bound)

Let $X_1, \dots, X_n$ be i.i.d. from $f(x; \theta)$ (under regularity conditions). For any unbiased estimator $\hat{\theta}$:

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n I(\theta)}$$

Back to the height example: We found $I(\mu) = 1/\sigma^2$, so:

$$\text{CRLB} = \frac{1}{n \cdot (1/\sigma^2)} = \frac{\sigma^2}{n}$$

Does $\bar{X}$ attain the bound?

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} = \text{CRLB} \quad \checkmark$$

Yes. The sample mean achieves the lowest possible variance. There is **no unbiased estimator of μ that does better**

Regularity note

The CRLB requires the support of f not to depend on θ . This holds for Normal, Bernoulli, Poisson, Exponential—but **not** for Uniform(0, θ), where the bound does not apply.

Efficiency: How Well Are We Using the Data?

Definition

The **efficiency** of an unbiased estimator $\hat{\theta}$ is:

$$\text{eff}(\hat{\theta}) = \frac{\text{CRLB}}{\text{Var}(\hat{\theta})} = \frac{1/[n I(\theta)]}{\text{Var}(\hat{\theta})} \in (0, 1]$$

An estimator with $\text{eff} = 1$ is **efficient**: it achieves the Cramér–Rao lower bound.

$\bar{X}$ for μ (Normal)

$$\text{eff}(\bar{X}) = \frac{\sigma^2/n}{\sigma^2/n} = 1$$

Uses **all** n observations.

100% efficient. ✓

$\hat{p}$ for p (Bernoulli)

$$\text{eff}(\hat{p}) = \frac{p(1-p)/n}{p(1-p)/n} = 1$$

Sample proportion.

100% efficient. ✓

X_1 alone for μ

$$\text{eff}(X_1) = \frac{\sigma^2/n}{\sigma^2} = \frac{1}{n}$$

Ignores $n-1$ observations.

At $n=100$: **only 1% efficient!**

The Key Idea for Sufficient Statistics: Conditioning Removes Noise

Suppose $\hat{\theta}$ is an unbiased estimator of θ , but it may use the data *inefficiently*.

Suppose U is a **sufficient statistic** for θ .

Define a new estimator:

$$\hat{\theta}^* = \mathbb{E}(\hat{\theta} \mid U)$$

Think of $\hat{\theta}$ as a **noisy signal**. Part of its randomness comes from information *already captured* by U , and part is pure noise irrelevant to θ . Conditioning on U **strips away the irrelevant noise**, keeping only what matters.

Analogy: U is like a high-resolution satellite photo. $\hat{\theta}$ is like asking a random passerby for directions. If you *average* the passerby's answers over everyone who saw the same satellite photo, you filter out idiosyncratic confusion.

Remember the Rao–Blackwell Theorem: Formal Statement

Theorem (Rao–Blackwell)

Let $\hat{\theta}$ be an unbiased estimator of θ with $\mathbb{E}(\hat{\theta}^2) < \infty$, and let U be a sufficient statistic for θ . Define

$$\hat{\theta}^* = \mathbb{E}(\hat{\theta} \mid U).$$

Then:

- 1 $\hat{\theta}^*$ is a function of U alone (it is a statistic).
- 2 $\mathbb{E}(\hat{\theta}^*) = \theta$ (still unbiased).
- 3 $\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$ for all θ , with equality iff $\hat{\theta}$ is already a function of U .

What This Means

You can **never do worse** by conditioning on a sufficient statistic. You will usually do **strictly better**. The Rao–Blackwell procedure is a variance-reduction machine.

Remember the Proof of the Rao–Blackwell Theorem

Part 1: $\hat{\theta}^*$ is a statistic.

Since U is sufficient, $\mathbb{E}(\hat{\theta} \mid U)$ does not depend on θ (by the sufficiency condition on the conditional distribution). So $\hat{\theta}^*$ is a function of U alone. ✓

Remember the Proof of the Rao–Blackwell Theorem

Part 1: $\hat{\theta}^*$ is a statistic.

Since U is sufficient, $\mathbb{E}(\hat{\theta} \mid U)$ does not depend on θ (by the sufficiency condition on the conditional distribution). So $\hat{\theta}^*$ is a function of U alone. ✓

Part 2: Unbiasedness (Law of Iterated Expectations).

$$\mathbb{E}(\hat{\theta}^*) = \mathbb{E}[\mathbb{E}(\hat{\theta} \mid U)] = \mathbb{E}(\hat{\theta}) = \theta \quad \checkmark$$

Remember the Proof of the Rao–Blackwell Theorem

Part 1: $\hat{\theta}^*$ is a statistic.

Since U is sufficient, $\mathbb{E}(\hat{\theta} \mid U)$ does not depend on θ (by the sufficiency condition on the conditional distribution). So $\hat{\theta}^*$ is a function of U alone. ✓

Part 2: Unbiasedness (Law of Iterated Expectations).

$$\mathbb{E}(\hat{\theta}^*) = \mathbb{E}[\mathbb{E}(\hat{\theta} \mid U)] = \mathbb{E}(\hat{\theta}) = \theta \quad \checkmark$$

Part 3: Variance reduction (Law of Total Variance / Eve's Law).

$$\begin{aligned} \text{Var}(\hat{\theta}) &= \text{Var}[\mathbb{E}(\hat{\theta} \mid U)] + \mathbb{E}[\text{Var}(\hat{\theta} \mid U)] \\ &= \text{Var}(\hat{\theta}^*) + \underbrace{\mathbb{E}[\text{Var}(\hat{\theta} \mid U)]}_{\geq 0} \end{aligned}$$

Therefore $\text{Var}(\hat{\theta}) \geq \text{Var}(\hat{\theta}^*)$, with equality iff $\text{Var}(\hat{\theta} \mid U) = 0$ a.s., i.e., $\hat{\theta}$ is already a function of U . ✓

The Variance Decomposition: A Closer Look

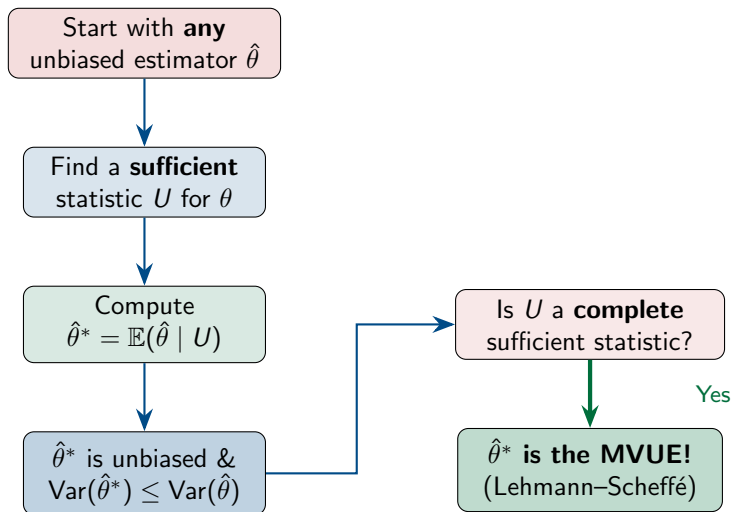
The key identity is the **Law of Total Variance**:

$$\underbrace{\text{Var}(\hat{\theta})}_{\text{total variance}} = \underbrace{\text{Var}[\mathbb{E}(\hat{\theta} \mid U)]}_{\text{"explained" variance}} + \underbrace{\mathbb{E}[\text{Var}(\hat{\theta} \mid U)]}_{\text{"unexplained" variance}}$$

- The **first term** is the variance of the Rao–Blackwellized estimator $\hat{\theta}^*$ —the useful part.
- The **second term** is the variance that comes from randomness *not captured* by U —pure noise.
- Conditioning on U **eliminates** the second term entirely.

Think of an **ANOVA decomposition**: the sufficient statistic explains the signal; the residual is noise. Rao–Blackwell says: *throw away the residual*.

The Rao–Blackwell Procedure: A Flowchart



The Minimum Variance Unbiased Estimator (MVUE)

Definition

An estimator $\hat{\theta}^*$ is the **Minimum Variance Unbiased Estimator (MVUE)** of θ if:

- 1 $\mathbb{E}(\hat{\theta}^*) = \theta$ (unbiased)
- 2 $\text{Var}(\hat{\theta}^*) \leq \text{Var}(\tilde{\theta})$ for **every** other unbiased estimator $\tilde{\theta}$, for all θ

The MVUE is the **most precise** unbiased estimator possible. No other unbiased estimator can beat it in terms of variability—it squeezes out every drop of information from the data.

It is the statistical analog of a **perfectly efficient market**: no information is wasted, no improvement is possible.

Does MVUE Always Exist?

- Not always, but for most standard families (exponential family) it does
- When it exists, it is **unique**

How to Find the MVUE: Two Strategies

Strategy 1: Rao–Blackwell + Lehmann–Scheffé

- 1 Find a **complete sufficient** statistic T
- 2 Find *any* unbiased estimator $\hat{\theta}$
- 3 Compute $\hat{\theta}^* = \mathbb{E}(\hat{\theta} \mid T)$
- 4 By Lehmann–Scheffé: $\hat{\theta}^*$ is the **MVUE**

Strategy 2: Direct approach

- 1 Find a complete sufficient statistic T
- 2 Find a function $\delta(T)$ such that $\mathbb{E}[\delta(T)] = \theta$
- 3 Then $\delta(T)$ is automatically the MVUE (no need to start with another estimator)

Strategy 2 is often easier in practice: if you can *directly* write an unbiased estimator as a function of a complete sufficient statistic, you are done!

Completeness: Why Do We Need It?

Rao–Blackwell guarantees *improvement*, but not that the result is the **best**. To guarantee the MVUE, we need one more ingredient: **completeness**.

Definition

A statistic T is **complete** for θ if:

$$\mathbb{E}[g(T)] = 0 \text{ for all } \theta \implies g(T) = 0 \text{ a.s.}$$

for any measurable function g .

Completeness says: the only unbiased estimator of zero that is a function of T is the trivial estimator $g(T) \equiv 0$.

In plain language: The only way a function of T can have expected value zero **for every parameter value** is if that function is actually **zero** (almost surely).

Completeness: Examples

So a complete statistic contains no redundant information that can “cancel out” across all parameter values.

Complete Sufficient Statistics

- Bernoulli(p): $\sum X_i$
- $N(\mu, \sigma_0^2)$ (σ_0^2 known): $\sum X_i$
- $N(\mu, \sigma^2)$ (both unknown):
($\sum X_i, \sum X_i^2$)
- Poisson(λ): $\sum X_i$
- Exponential(θ): $\sum X_i$

General rule: For the **exponential family**, the natural sufficient statistic is complete.

Why Completeness Matters

Without completeness, two different functions of T could both be unbiased for θ —and we could not declare either one the MVUE.

Completeness **ensures uniqueness**: there is exactly *one* unbiased estimator based on T .

The Lehmann–Scheffé Theorem

Theorem (Lehmann–Scheffé)

Let T be a complete sufficient statistic for θ . If $\delta(T)$ is any unbiased estimator of θ that is a function of T , then $\delta(T)$ is the unique MVUE of θ .

Why does this work?

- By Rao–Blackwell, any unbiased estimator can be improved by conditioning on T .
- After conditioning, the result is a function of T .
- By completeness, there is only **one** unbiased function of T —so it must be the MVUE.

Lehmann–Scheffé is the “final seal”: once you have a complete sufficient statistic and an unbiased function of it, you have found the **best possible estimator**. The search is over.

Example 1: Bernoulli MVUE for p

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$ is complete and sufficient (exponential family).

Example 1: Bernoulli MVUE for p

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$ is complete and sufficient (exponential family).

Step 2: Find an unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{np}{n} = p \quad \checkmark$$

Example 1: Bernoulli MVUE for p

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i \sim \text{Binomial}(n, p)$ is complete and sufficient (exponential family).

Step 2: Find an unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{np}{n} = p \quad \checkmark$$

Conclusion: By Lehmann–Scheffé, $\hat{p}_{\text{MVUE}} = \bar{X} = T/n$ is the **MVUE** for p .

Its variance is $\text{Var}(\bar{X}) = \frac{p(1-p)}{n}$.

This also attains the **Cramér–Rao lower bound** (check!), confirming it is the best we can do.

Example 2: Normal Distribution — MVUE for μ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ with σ^2 known.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient (exponential family).

Example 2: Normal Distribution — MVUE for μ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ with σ^2 known.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient (exponential family).

Step 2: Unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{n\mu}{n} = \mu \quad \checkmark$$

Example 2: Normal Distribution — MVUE for μ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$ with σ^2 known.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient (exponential family).

Step 2: Unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{n\mu}{n} = \mu \quad \checkmark$$

Conclusion: $\hat{\mu}_{\text{MVUE}} = \bar{X}$ with $\text{Var}(\bar{X}) = \sigma^2/n$.

When both μ and σ^2 are unknown:

- Complete sufficient statistic: $(\sum X_i, \sum X_i^2)$
- MVUE for μ : $\bar{X}$
- MVUE for σ^2 : $S^2 = \frac{1}{n-1} \sum (X_i - \bar{X})^2$

Example 3: Exponential Distribution — MVUE for θ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\theta)$ with density $f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$, $x > 0$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient.

Note: $T \sim \text{Gamma}(n, \theta)$, so $\mathbb{E}(T) = n\theta$.

Example 3: Exponential Distribution — MVUE for θ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\theta)$ with density $f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$, $x > 0$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient.

Note: $T \sim \text{Gamma}(n, \theta)$, so $\mathbb{E}(T) = n\theta$.

Step 2: Unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{n\theta}{n} = \theta \quad \checkmark$$

Example 3: Exponential Distribution — MVUE for θ

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\theta)$ with density $f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}$, $x > 0$.

Step 1: Complete sufficient statistic.

$T = \sum_{i=1}^n X_i$ is complete and sufficient.

Note: $T \sim \text{Gamma}(n, \theta)$, so $\mathbb{E}(T) = n\theta$.

Step 2: Unbiased function of T .

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{n\theta}{n} = \theta \quad \checkmark$$

Conclusion: $\hat{\theta}_{\text{MVUE}} = \bar{X}$ with $\text{Var}(\bar{X}) = \theta^2/n$.

A recurring pattern: for many one-parameter exponential family distributions, the **sample mean is the MVUE**. The sufficient statistic is the sum, and dividing by n gives unbiasedness.

Example 4: Weibull Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Weibull}(\beta, \theta)$ with known shape β and density:

$$f(x; \theta) = \frac{\beta}{\theta} x^{\beta-1} e^{-x^\beta/\theta}, \quad x > 0$$

Step 1: Sufficient statistic.

The joint density is $\prod_{i=1}^n \frac{\beta}{\theta} x_i^{\beta-1} e^{-x_i^\beta/\theta} = \frac{\beta^n}{\theta^n} \left(\prod x_i^{\beta-1} \right) \exp\left(-\frac{\sum x_i^\beta}{\theta}\right)$

By the Factorization Theorem: $T = \sum_{i=1}^n X_i^\beta$ is sufficient (and complete).

Example 4: Weibull Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Weibull}(\beta, \theta)$ with known shape β and density:

$$f(x; \theta) = \frac{\beta}{\theta} x^{\beta-1} e^{-x^\beta/\theta}, \quad x > 0$$

Step 1: Sufficient statistic.

The joint density is $\prod_{i=1}^n \frac{\beta}{\theta} x_i^{\beta-1} e^{-x_i^\beta/\theta} = \frac{\beta^n}{\theta^n} \left(\prod x_i^{\beta-1} \right) \exp\left(-\frac{\sum x_i^\beta}{\theta}\right)$

By the Factorization Theorem: $T = \sum_{i=1}^n X_i^\beta$ is sufficient (and complete).

Step 2: Note that $Y_i = X_i^\beta \sim \text{Exp}(\theta)$, so $T = \sum Y_i \sim \text{Gamma}(n, \theta)$.
Therefore $\mathbb{E}(T/n) = \theta$.

Conclusion: $\hat{\theta}_{\text{MVUE}} = \frac{1}{n} \sum_{i=1}^n X_i^\beta$.

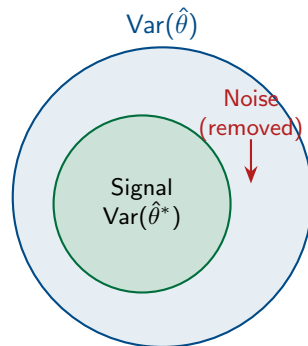
The Weibull reduces to the Exponential case via the transformation $Y_i = X_i^\beta$. Recognizing such transformations is a powerful technique!

The sufficient statistic acts as a **noise filter**. It preserves all information about θ while discarding irrelevant randomness. Rao–Blackwell is essentially a **denoising** operation.

Consider all the randomness in $\hat{\theta}$. It comes from two sources:

- 1 **Signal:** Variation that is *related to* θ —this is captured by the sufficient statistic U .
- 2 **Noise:** Variation that is *unrelated to* θ —this is the randomness in $\hat{\theta}$ *given* U .

Conditioning on U eliminates source (2) entirely.



Analogy: The Wisdom of the “Right” Crowd

Imagine estimating the number of marbles in a jar:

- **Before Rao–Blackwell:** Ask one random person (X_1). Their guess is unbiased but very noisy.
- **After Rao–Blackwell:** Average the guesses of *everyone who saw the same evidence* (condition on sufficient statistic). Individual quirks cancel out.

Key Insight

The sufficient statistic U determines **what evidence matters**. Conditioning on U is like asking: “Among all estimators that saw the same relevant evidence, what is the consensus?”

This consensus always has **less variance** than any individual guess.

The Complete MVUE Recipe

1. Write the joint density/pmf



2. Apply **Factorization Theorem** to find sufficient statistic T

Weeks 3–4:
Sufficiency via
Factorization



3. Verify T is **complete** (exponential family $\Rightarrow$ automatic)



4. Find a function $\delta(T)$ with $\mathbb{E}[\delta(T)] = \theta$

Often: just divide
the sufficient stat
by the right
constant



5. By Lehmann–Scheffé:
 $\delta(T)$ is the **MVUE**

Problem 1: Poisson MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$. Find the MVUE for λ .

Problem 1: Poisson MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$. Find the MVUE for λ .

Solution:

Step 1: The Poisson family is an exponential family. The sufficient statistic is

$$T = \sum_{i=1}^n X_i \sim \text{Poisson}(n\lambda).$$

Since Poisson is a full exponential family, T is **complete**.

Step 2: Find an unbiased function of T :

$$\mathbb{E}\left(\frac{T}{n}\right) = \frac{n\lambda}{n} = \lambda \quad \checkmark$$

Conclusion: By Lehmann–Scheffé, $\hat{\lambda}_{\text{MVUE}} = \bar{X} = T/n$.

$\text{Var}(\bar{X}) = \lambda/n$, which equals the CRLB (verify: $I(\lambda) = 1/\lambda$, so $\text{CRLB} = \lambda/n$). $\checkmark$

Problem 2: Uniform MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find the MVUE for θ .

Problem 2: Uniform MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find the MVUE for θ .

Solution:

Step 1: The sufficient statistic is $T = X_{(n)} = \max(X_1, \dots, X_n)$. Its CDF is $F_T(t) = (t/\theta)^n$ for $0 < t < \theta$, so $\mathbb{E}(T) = \frac{n}{n+1}\theta$.

T is complete and sufficient.

Problem 2: Uniform MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find the MVUE for θ .

Solution:

Step 1: The sufficient statistic is $T = X_{(n)} = \max(X_1, \dots, X_n)$. Its CDF is $F_T(t) = (t/\theta)^n$ for $0 < t < \theta$, so $\mathbb{E}(T) = \frac{n}{n+1}\theta$.

T is complete and sufficient.

Step 2: $\mathbb{E}(T) = \frac{n}{n+1}\theta \neq \theta$, so T itself is biased. Correct it:

$$\delta(T) = \frac{n+1}{n} T = \frac{n+1}{n} X_{(n)}$$

Then $\mathbb{E}[\delta(T)] = \frac{n+1}{n} \cdot \frac{n}{n+1}\theta = \theta \checkmark$

Conclusion: $\hat{\theta}_{\text{MVUE}} = \frac{n+1}{n} X_{(n)}$.

Here $\bar{X}$ is also unbiased, but it is **not** the MVUE! The order statistic $X_{(n)}$ captures the sufficient information, and $\bar{X}$ is not a function of $X_{(n)}$.

Problem 3: Normal Variance MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both parameters unknown. Show that S^2 is the MVUE for σ^2 .

Problem 3: Normal Variance MVUE

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both parameters unknown. Show that S^2 is the MVUE for σ^2 .

Solution:

Step 1: The complete sufficient statistic is $(\bar{X}, \sum (X_i - \bar{X})^2)$.

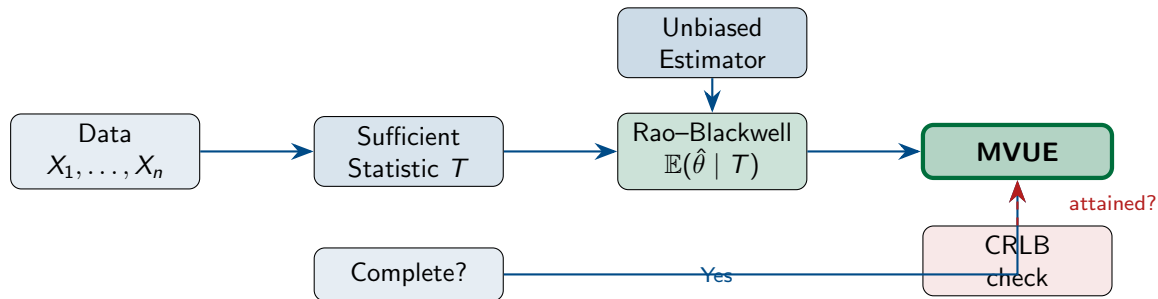
We know that $\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$, so $\mathbb{E}[(n-1)S^2/\sigma^2] = n-1$.

Step 2: Therefore $\mathbb{E}(S^2) = \sigma^2$, and S^2 is a function of $\sum (X_i - \bar{X})^2$, which is part of the complete sufficient statistic.

Conclusion: By Lehmann–Scheffé, $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ is the MVUE for σ^2 .

This is exactly why we divide by $n-1$ and not n ! The factor $n-1$ makes S^2 unbiased, and being a function of the complete sufficient statistic guarantees it is the **best** we can do.

The Big Picture: Estimation Theory Map



The whole theory fits together beautifully:

- **Sufficiency** captures all information about θ
- **Rao-Blackwell** uses that information optimally (removes noise)
- **Completeness + Lehmann-Scheffé** guarantees uniqueness

Key Takeaways from Week 5

- 1 The **Rao–Blackwell Theorem** says: conditioning an unbiased estimator on a sufficient statistic **never increases variance** and usually **strictly reduces** it
- 2 The proof uses two fundamental tools: the **Law of Iterated Expectations** (preserves unbiasedness) and the **Law of Total Variance** (shows variance reduction)
- 3 The **MVUE** is the unbiased estimator with the smallest possible variance—the best we can achieve without introducing bias
- 4 **Lehmann–Scheffé**: if T is complete sufficient and $\delta(T)$ is unbiased, then $\delta(T)$ is the MVUE—**unique and optimal**
- 5 **Practical recipe**: find complete sufficient statistic $\rightarrow$ construct an unbiased function of it $\rightarrow$ done!
- 6 For exponential family distributions, the MVUE typically **attains the CRLB**

Coming Up Next Week

Week 6 Preview

Methods of Estimation: Maximum Likelihood Revisited

- Large-sample properties of the MLE: consistency, asymptotic normality, asymptotic efficiency
- The MLE as an approximately MVUE in large samples
- Delta method and variance-stabilizing transformations
- When exact MVUE is hard to find, the MLE is our best friend

The MVUE theory we learned today is “exact” and beautiful, but requires knowing the complete sufficient statistic. In practice, the MLE is often easier to compute and performs nearly as well in large samples. Next week bridges the gap between exact and asymptotic optimality.

- Wackerly, Mendenhall & Scheaffer, **Ch. 9**: Properties of Point Estimators and Methods of Estimation
 - § 9.4: The Rao–Blackwell Theorem
 - § 9.5: Minimum-Variance Unbiased Estimation
- Hogg, McKean & Craig, **Ch. 7**: Sufficiency, Completeness, and MVUE
 - § 7.3: Completeness and Uniqueness
 - § 7.4: The Rao–Blackwell Theorem
- Casella & Berger, **Ch. 7**: Point Estimation (for a more advanced treatment)