

Statistics for Economists II

Week 4: Sufficient Statistics & the Cramér–Rao Lower Bound

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Motivation: Data Compression Without Information Loss
- 2 Sufficient Statistics: Definition and Theory
- 3 Examples: Finding Sufficient Statistics
- 4 Examples: Finding Sufficient Statistics
- 5 The Rao–Blackwell Theorem
- 6 Minimum-Variance Unbiased Estimation (MVUE)
- 7 MVUE Examples
- 8 Connecting the Pieces
- 9 Key Takeaways

The Big Question

So far we learned about point estimators and desirable properties. Today we ask two deeper questions:

- 1 **Data reduction:** Can we *compress* the sample without losing any information about θ ?

Economics Motivation

Central banks collect *millions* of data points. Can they summarize all that data into a few key numbers without losing information?

And when they estimate inflation, **what is the minimum uncertainty they must live with?**

The Big Question

So far we learned about point estimators and desirable properties. Today we ask two deeper questions:

- 1 **Data reduction:** Can we *compress* the sample without losing any information about θ ?
- 2 **Fundamental limits:** Is there a *minimum possible variance* for any unbiased estimator?

Economics Motivation

Central banks collect *millions* of data points. Can they summarize all that data into a few key numbers without losing information?

And when they estimate inflation, **what is the minimum uncertainty they must live with?**

The Big Question

So far we learned about point estimators and desirable properties. Today we ask two deeper questions:

- 1 **Data reduction:** Can we *compress* the sample without losing any information about θ ?
- 2 **Fundamental limits:** Is there a *minimum possible variance* for any unbiased estimator?

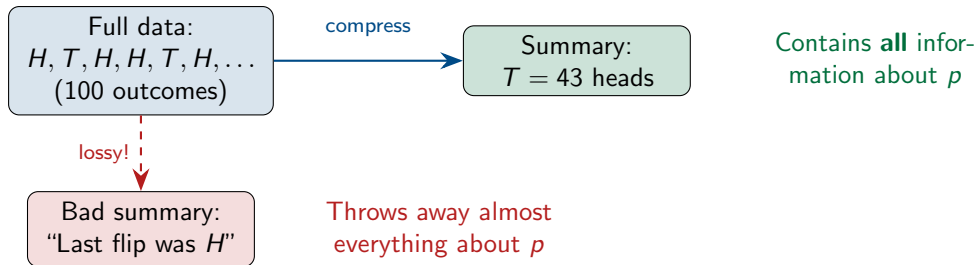
These two questions lead to **sufficient statistics** and the **Cramér–Rao inequality**.

Economics Motivation

Central banks collect *millions* of data points. Can they summarize all that data into a few key numbers without losing information?

And when they estimate inflation, **what is the minimum uncertainty they must live with?**

The Coin Flip Story



You flip a coin 100 times to estimate $p = \mathbb{P}(\text{heads})$. Your assistant records the outcome of every single flip. But you only need to know **how many heads**: $T = \sum X_i = 43$. This single number contains **all the information** about p that was in the original 100 observations. We say T is a **sufficient statistic** for p .

Why Does Only the Count Matter?

Consider two scenarios. Each has 100 coin flips with $T = 43$ heads:

Sequence A: $H, T, H, H, T, T, H, T, H, H, T, \dots$ (43 heads, 57 tails)

Sequence B: $T, H, T, H, H, H, T, H, T, T, H, \dots$ (43 heads, 57 tails)

Why Does Only the Count Matter?

Consider two scenarios. Each has 100 coin flips with $T = 43$ heads:

Sequence A: $H, T, H, H, T, T, H, T, H, H, T, \dots$ (43 heads, 57 tails)

Sequence B: $T, H, T, H, H, H, T, H, T, T, H, \dots$ (43 heads, 57 tails)

The likelihood of *any* specific sequence with exactly k heads is:

$$L(p) = p^k(1 - p)^{n-k}$$

The **order doesn't matter**—only k does! Given $T = 43$, the conditional distribution of the data does not depend on p at all.

Key Insight

Once you know $T = \sum X_i$, the remaining information in the sample (the ordering) is **pure noise**—it tells you nothing about p . This is exactly what sufficiency means.

Why Does Only the Count Matter?

The key distinction is between the likelihood of a specific observed sequence vs. the probability of a count. When you observe a specific sequence of coin flips — say, HHTHT — the probability of that exact sequence is simply:

$$L(p) = p \cdot p \cdot (1 - p) \cdot p \cdot (1 - p) = p^3(1 - p)^2$$

No combinations needed, because the order is already fixed — you observed that particular sequence. The likelihood function is defined for the data you actually observed. The binomial coefficient $\binom{n}{k}$ only enters when you ask: "What is the probability of getting k heads in any order?" That's the PMF of the Binomial distribution:

$$P(T = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

In the likelihood context, you're not summing over all possible orderings — you're evaluating the probability of the one specific sample you got.

Why Does Only the Count Matter?

And here's the beautiful sufficiency point:

Every possible sequence with exactly k heads produces the same likelihood $p^k(1 - p)^{n-k}$, regardless of where the heads appeared. So the specific ordering is irrelevant — only the count k matters for inference about p . That's precisely why $T = \sum X_i$ is a sufficient statistic. Even if you did include the $\binom{n}{k}$ (as a proportionality constant), it wouldn't change anything — it's a constant that doesn't depend on p .

Formal Definition of Sufficiency

Definition (Sufficient Statistic)

A statistic $T = T(X_1, \dots, X_n)$ is **sufficient** for θ if the conditional distribution of the sample $X = (X_1, \dots, X_n)$ given $T = t$ does not depend on θ .

Formal Definition of Sufficiency

Definition (Sufficient Statistic)

A statistic $T = T(X_1, \dots, X_n)$ is **sufficient** for θ if the conditional distribution of the sample $X = (X_1, \dots, X_n)$ given $T = t$ does not depend on θ .

Sufficiency means **data compression without information loss**:

- If T is sufficient, then once you know T , the rest of the data is “uninformative noise”
- The conditional distribution $f(x \mid T = t)$ is free of θ —you cannot learn anything more about θ by looking at the raw data beyond what T already tells you
- Any inference about θ based on the full data can be *equally well* based on T alone

Formal Definition of Sufficiency

Definition (Sufficient Statistic)

A statistic $T = T(X_1, \dots, X_n)$ is **sufficient** for θ if the conditional distribution of the sample $X = (X_1, \dots, X_n)$ given $T = t$ does not depend on θ .

Sufficiency means **data compression without information loss**:

- If T is sufficient, then once you know T , the rest of the data is “uninformative noise”
- The conditional distribution $f(x \mid T = t)$ is free of θ —you cannot learn anything more about θ by looking at the raw data beyond what T already tells you
- Any inference about θ based on the full data can be *equally well* based on T alone

Important Note

Sufficient statistics are **not unique**. The entire sample $(X_1, \dots, X_n)$ is always (trivially) sufficient. The goal is to find the *simplest* sufficient statistic.

The Factorization Theorem (Fisher–Neyman)

Theorem (Factorization Theorem)

Let $f(x|\theta)$ denote the joint pdf (or pmf) of the sample. Then $T(x)$ is a sufficient statistic for θ if and only if there exist functions g and h such that:

$$f(x|\theta) = g(T(x), \theta) \cdot h(x)$$

where g depends on x only through $T(x)$, and h does not depend on θ .

The Factorization Theorem (Fisher–Neyman)

Theorem (Factorization Theorem)

Let $f(x|\theta)$ denote the joint pdf (or pmf) of the sample. Then $T(x)$ is a sufficient statistic for θ if and only if there exist functions g and h such that:

$$f(x|\theta) = g(T(x), \theta) \cdot h(x)$$

where g depends on x only through $T(x)$, and h does not depend on θ .

This theorem gives us a **practical recipe** for finding sufficient statistics:

- 1 Write down the joint likelihood $f(x|\theta)$
- 2 Try to split it into two pieces: one involving θ and the data only through T , and a leftover piece that is free of θ
- 3 If you can do this, T is sufficient!

Example 1: Normal Distribution (Known Variance)

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

Example 1: Normal Distribution (Known Variance)

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

Solution: The joint pdf is:

$$f(\mathbf{x}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

Example 1: Normal Distribution (Known Variance)

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

Solution: The joint pdf is:

$$f(\mathbf{x}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

Expand the exponent:

$$\sum (x_i - \mu)^2 = \sum x_i^2 - 2\mu \sum x_i + n\mu^2$$

So we can write:

$$f(\mathbf{x}|\mu) = \underbrace{\exp\left(-\frac{1}{2\sigma^2} [-2\mu \cdot n\bar{x} + n\mu^2]\right)}_{g(\bar{x}, \mu)} \cdot (2\pi\sigma^2)^{-n/2} \cdot \underbrace{\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)}_{h(\mathbf{x})}$$

Example 1: Normal Distribution (Known Variance)

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

Solution: The joint pdf is:

$$f(\mathbf{x}|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x_i - \mu)^2}{2\sigma^2}\right) = \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2\right)$$

Expand the exponent:

$$\sum (x_i - \mu)^2 = \sum x_i^2 - 2\mu \sum x_i + n\mu^2$$

So we can write:

$$f(\mathbf{x}|\mu) = \underbrace{\exp\left(-\frac{1}{2\sigma^2} [-2\mu \cdot n\bar{x} + n\mu^2]\right)}_{g(\bar{x}, \mu)} \cdot (2\pi\sigma^2)^{-n/2} \cdot \underbrace{\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)}_{h(\mathbf{x})}$$

By Factorization Theorem:

$$T = \bar{X} = \frac{1}{n} \sum X_i \text{ is sufficient for } \mu$$

Example 1: Normal Distribution (Known Variance)

The Factorization Theorem (Fisher-Neyman) says: T is sufficient for θ if and only if you can write:

$$f(x|\theta) = g(T(x), \theta) \cdot h(x)$$

where:

- g depends on the data only through $T(x)$ — and also on θ
- h depends on the data but NOT on θ

Now look at your factorization:

- The g part: $\exp\left(-\frac{1}{2\sigma^2} \left[-2\mu \cdot n\bar{x} + n\mu^2\right]\right) \cdot (2\pi\sigma^2)^{-n/2}$
- Ask yourself: how does this piece depend on the data? The only data-dependent term in here is $n\bar{x}$. Since n is a known constant, knowing $n\bar{x}$ is the same as knowing $\bar{x}$. So g depends on the data only through $\bar{x}$. It also depends on μ

Example 1: Normal Distribution (Known Variance)

The h part: $\exp\left(-\frac{\sum x_i^2}{2\sigma^2}\right)$

This depends on the individual x_i 's, but — crucially — there is no μ anywhere in this term. (Remember, σ^2 is treated as a known constant in this problem.)

So the conclusion follows directly: since we managed to split the joint density into a piece that touches the data only via $\bar{x}$ (and contains μ) times a piece that involves the raw data but not μ , the Factorization Theorem tells us $T = \bar{X}$ is sufficient for μ .

One small note: Technically, $T = \sum X_i$ or $T = n\bar{X}$ would be equally valid sufficient statistics — they're all one-to-one functions of each other (since n is fixed). We write $\bar{X}$ because it's the most natural and interpretable form.

Example 2: Bernoulli Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$. Find a sufficient statistic for p .

Example 2: Bernoulli Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$. Find a sufficient statistic for p .

Solution: The joint pmf is:

$$f(\mathbf{x}|p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

Example 2: Bernoulli Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$. Find a sufficient statistic for p .

Solution: The joint pmf is:

$$f(\mathbf{x}|p) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^{\sum x_i} (1-p)^{n-\sum x_i}$$

This is already factored! Set $T = \sum_{i=1}^n x_i$:

$$f(\mathbf{x}|p) = \underbrace{p^T (1-p)^{n-T}}_{g(T, p)} \cdot \underbrace{1}_{h(\mathbf{x})}$$

By Factorization Theorem:

$$T = \sum_{i=1}^n X_i \text{ is sufficient for } p$$

This is exactly our coin flip story! The total number of heads $T = \sum X_i$ captures everything about p . The order of heads and tails is irrelevant—it lives in $h(\mathbf{x}) = 1$.

Example 3: Uniform Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find a sufficient statistic for θ .

Example 3: Uniform Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find a sufficient statistic for θ .

Solution: The joint pdf is:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{1}{\theta} 1(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} 1(0 \leq x_{(1)}) \cdot 1(x_{(n)} \leq \theta)$$

where $x_{(1)} = \min_i x_i$ and $x_{(n)} = \max_i x_i$.

Example 3: Uniform Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find a sufficient statistic for θ .

Solution: The joint pdf is:

$$f(\mathbf{x}|\theta) = \prod_{i=1}^n \frac{1}{\theta} 1(0 \leq x_i \leq \theta) = \frac{1}{\theta^n} 1(0 \leq x_{(1)}) \cdot 1(x_{(n)} \leq \theta)$$

where $x_{(1)} = \min_i x_i$ and $x_{(n)} = \max_i x_i$.

$$f(\mathbf{x}|\theta) = \underbrace{\frac{1}{\theta^n} 1(x_{(n)} \leq \theta)}_{g(x_{(n)}, \theta)} \cdot \underbrace{1(x_{(1)} \geq 0)}_{h(\mathbf{x})}$$

By Factorization Theorem: $T = X_{(n)} = \max(X_1, \dots, X_n)$ is sufficient for θ

For the uniform distribution, only the **maximum observation** matters. It makes sense: the maximum tells us the most about where the upper boundary θ is. All other observations just confirm they are below θ .

Summary of Sufficiency Examples

Distribution	Parameter	Sufficient Statistic
$N(\mu, \sigma_0^2), \sigma_0^2 \text{ known}$	μ	$\bar{X}$
Bernoulli(p)	p	$\sum X_i$
Poisson(λ)	λ	$\sum X_i$
Uniform($0, \theta$)	θ	$X_{(n)} = \max X_i$
Uniform(θ_1, θ_2)	(θ_1, θ_2)	$(X_{(1)}, X_{(n)})$
$N(\mu, \sigma^2), \text{ both unknown}$	(μ, σ^2)	$(\sum X_i, \sum X_i^2)$
Exponential(β)	β	$\sum X_i$

Notice the pattern: for **exponential family** distributions (Normal, Bernoulli, Poisson, Exponential, etc.), sufficient statistics are sums or simple functions of the data. For non-exponential-family distributions like the Uniform, the sufficient statistic involves **order statistics**.

Motivation: Can We Improve Any Estimator?

We learned that **sufficient statistics** compress the data without losing information about θ .

Now the natural question:

The Key Question

If we have an unbiased estimator $\hat{\theta}$, can we **improve** it (reduce its variance) by using a sufficient statistic?

Motivation: Can We Improve Any Estimator?

We learned that **sufficient statistics** compress the data without losing information about θ .

Now the natural question:

The Key Question

If we have an unbiased estimator $\hat{\theta}$, can we **improve** it (reduce its variance) by using a sufficient statistic?

Answer: Yes! The **Rao–Blackwell theorem** tells us exactly how.

Imagine two economists estimating average income:

- Economist A uses *some* of the survey data
- Economist B uses *all* the information (the sufficient statistic)

Economist B can **never do worse**—and will usually do **strictly better**.

A Quick Recall: Two Useful Theorems

Before stating Rao–Blackwell, recall two results about conditional expectation:

A Quick Recall: Two Useful Theorems

Before stating Rao–Blackwell, recall two results about conditional expectation:

Law of Iterated Expectations

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$. The overall mean equals the mean of the conditional means.

A Quick Recall: Two Useful Theorems

Before stating Rao–Blackwell, recall two results about conditional expectation:

Law of Iterated Expectations

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]$. The overall mean equals the mean of the conditional means.

Law of Total Variance

$$\text{Var}(X) = \underbrace{\text{Var}(\mathbb{E}[X | Y])}_{\text{variance of conditional means}} + \underbrace{\mathbb{E}[\text{Var}(X | Y)]}_{\text{average conditional variance} \geq 0}$$

A Quick Recall: Two Useful Theorems

Before stating Rao–Blackwell, recall two results about conditional expectation:

Law of Iterated Expectations

$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X | Y]]$. The overall mean equals the mean of the conditional means.

Law of Total Variance

$$\text{Var}(X) = \underbrace{\text{Var}(\mathbb{E}[X | Y])}_{\text{variance of conditional means}} + \underbrace{\mathbb{E}[\text{Var}(X | Y)]}_{\text{average conditional variance} \geq 0}$$

The second term is always ≥ 0 , so $\text{Var}(\mathbb{E}[X | Y]) \leq \text{Var}(X)$. Conditioning **reduces** (or at least does not increase) variance!

The Rao–Blackwell Theorem

Theorem (Rao–Blackwell Theorem)

Let $\hat{\theta}$ be an unbiased estimator of θ with $\text{Var}(\hat{\theta}) < \infty$. Let U be a sufficient statistic for θ . Define a new estimator: $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid U]$. Then, for all θ :

- 1 $\mathbb{E}[\hat{\theta}^*] = \theta$ (still unbiased)
- 2 $\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$ (variance is no larger)

The Rao–Blackwell Theorem

Theorem (Rao–Blackwell Theorem)

Let $\hat{\theta}$ be an unbiased estimator of θ with $\text{Var}(\hat{\theta}) < \infty$. Let U be a sufficient statistic for θ . Define a new estimator: $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid U]$. Then, for all θ :

- ① $\mathbb{E}[\hat{\theta}^*] = \theta$ (still unbiased)
- ② $\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta})$ (variance is no larger)

The Rao–Blackwell theorem gives a **recipe for improvement**: take *any* unbiased estimator, condition it on a sufficient statistic, and you get an estimator that is:

- still unbiased,
- based only on the sufficient statistic,
- at least as good (and usually strictly better).

Proof of the Rao–Blackwell Theorem

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

By the law of total expectation: $\mathbb{E}[\hat{\theta}^*] = \mathbb{E}[\mathbb{E}[\hat{\theta} \mid U]] = \mathbb{E}[\hat{\theta}] = \theta \quad \checkmark$

Proof of the Rao–Blackwell Theorem

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

By the law of total expectation: $\mathbb{E}[\hat{\theta}^*] = \mathbb{E}[\mathbb{E}[\hat{\theta} \mid U]] = \mathbb{E}[\hat{\theta}] = \theta \quad \checkmark$

Part 2: $\hat{\theta}^*$ is a proper statistic.

Because U is sufficient, the conditional distribution of the data given U does **not depend on** θ . So $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid U]$ is a function of U alone. $\checkmark$

Proof of the Rao–Blackwell Theorem

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

By the law of total expectation: $\mathbb{E}[\hat{\theta}^*] = \mathbb{E}[\mathbb{E}[\hat{\theta} \mid U]] = \mathbb{E}[\hat{\theta}] = \theta \quad \checkmark$

Part 2: $\hat{\theta}^*$ is a proper statistic.

Because U is sufficient, the conditional distribution of the data given U does **not depend on** θ . So $\hat{\theta}^* = \mathbb{E}[\hat{\theta} \mid U]$ is a function of U alone. $\checkmark$

Part 3: Variance is reduced.

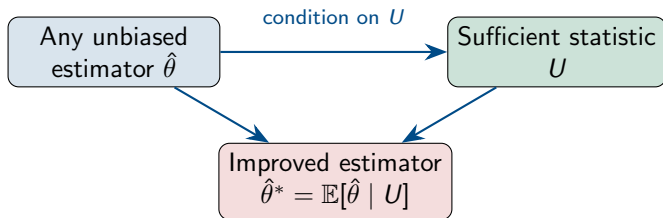
By the law of total variance:

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

Since the second term is ≥ 0 , we get $\text{Var}(\hat{\theta}) \geq \text{Var}(\hat{\theta}^*)$. $\checkmark$

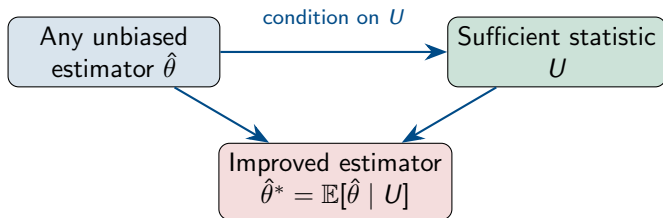


What Does Rao–Blackwell Really Say?



Still unbiased, smaller (or equal) variance

What Does Rao–Blackwell Really Say?



Still unbiased, smaller (or equal) variance

Important Remarks

- Reapplying Rao–Blackwell with the **same** U gives **no further improvement** (because $\hat{\theta}^*$ is already a function of U).
- So use the **best** sufficient statistic from the start—the one from the factorization criterion.
- In practice: directly look for an unbiased function of the sufficient statistic.

What Is an MVUE?

Definition (Minimum-Variance Unbiased Estimator)

An estimator $\hat{\theta}$ is called a **minimum-variance unbiased estimator (MVUE)** of θ if:

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

What Is an MVUE?

Definition (Minimum-Variance Unbiased Estimator)

An estimator $\hat{\theta}$ is called a **minimum-variance unbiased estimator (MVUE)** of θ if:

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

An MVUE is the “gold standard” among unbiased estimators: it hits the target on average (unbiased), has the **tightest possible spread**, and no other unbiased estimator can beat it!

What Is an MVUE?

Definition (Minimum-Variance Unbiased Estimator)

An estimator $\hat{\theta}$ is called a **minimum-variance unbiased estimator (MVUE)** of θ if:

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

An MVUE is the “gold standard” among unbiased estimators: it hits the target on average (unbiased), has the **tightest possible spread**, and no other unbiased estimator can beat it!

The Rao–Blackwell Connection

The Rao–Blackwell theorem tells us: **always look among functions of sufficient statistics**. If we find an unbiased function of the sufficient statistic, we have our MVUE.

The MVUE Recipe

Two-Step Procedure for Finding MVUEs

Step 1: Find the sufficient statistic U using the **factorization criterion**.

Step 2: Find a function $h(U)$ such that $\mathbb{E}[h(U)] = \theta$.

Then $h(U)$ is the MVUE for θ .

The MVUE Recipe

Two-Step Procedure for Finding MVUEs

Step 1: Find the sufficient statistic U using the **factorization criterion**.

Step 2: Find a function $h(U)$ such that $\mathbb{E}[h(U)] = \theta$.

Then $h(U)$ is the MVUE for θ .

Why does this work?

- By Step 1, U captures **all** the information about θ .
- By Step 2, $h(U)$ is unbiased.
- By Rao–Blackwell, no estimator that ignores the sufficient statistic can do better.
- Since $h(U)$ is already a function of U , it cannot be improved further.

The MVUE Recipe

Two-Step Procedure for Finding MVUEs

Step 1: Find the sufficient statistic U using the **factorization criterion**.

Step 2: Find a function $h(U)$ such that $\mathbb{E}[h(U)] = \theta$.

Then $h(U)$ is the MVUE for θ .

Why does this work?

- By Step 1, U captures **all** the information about θ .
- By Step 2, $h(U)$ is unbiased.
- By Rao–Blackwell, no estimator that ignores the sufficient statistic can do better.
- Since $h(U)$ is already a function of U , it cannot be improved further.

Let's see this recipe in action with several examples.

Example 1: Bernoulli — MVUE for p

Setup: $Y_1, \dots, Y_n$ are i.i.d. Bernoulli(p), where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$.

Example 1: Bernoulli — MVUE for p

Setup: $Y_1, \dots, Y_n$ are i.i.d. Bernoulli(p), where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$.

Step 1: Find the sufficient statistic.

The likelihood is:

$$L(p) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} = \underbrace{p^{\sum y_i} (1-p)^{n-\sum y_i}}_{g(\sum y_i, p)} \cdot \underbrace{1}_{h(y_1, \dots, y_n)}$$

By the factorization criterion: $U = \sum_{i=1}^n Y_i$ is sufficient for p .

Example 1: Bernoulli — MVUE for p

Setup: $Y_1, \dots, Y_n$ are i.i.d. Bernoulli(p), where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$.

Step 1: Find the sufficient statistic.

The likelihood is:

$$L(p) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} = \underbrace{p^{\sum y_i} (1-p)^{n-\sum y_i}}_{g(\sum y_i, p)} \cdot \underbrace{1}_{h(y_1, \dots, y_n)}$$

By the factorization criterion: $U = \sum_{i=1}^n Y_i$ is sufficient for p .

Step 2: Find an unbiased function of U .

We need $h(U)$ with $\mathbb{E}[h(U)] = p$. Since $\mathbb{E}[U] = np$:

$$\mathbb{E}\left[\frac{U}{n}\right] = \frac{np}{n} = p \quad \checkmark$$

Example 1: Bernoulli — MVUE for p

Setup: $Y_1, \dots, Y_n$ are i.i.d. Bernoulli(p), where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$.

Step 1: Find the sufficient statistic.

The likelihood is:

$$L(p) = \prod_{i=1}^n p^{y_i} (1-p)^{1-y_i} = \underbrace{p^{\sum y_i} (1-p)^{n-\sum y_i}}_{g(\sum y_i, p)} \cdot \underbrace{1}_{h(y_1, \dots, y_n)}$$

By the factorization criterion: $U = \sum_{i=1}^n Y_i$ is sufficient for p .

Step 2: Find an unbiased function of U .

We need $h(U)$ with $\mathbb{E}[h(U)] = p$. Since $\mathbb{E}[U] = np$:

$$\mathbb{E}\left[\frac{U}{n}\right] = \frac{np}{n} = p \quad \checkmark$$

Conclusion: $\hat{p} = \bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$ is the **MVUE** for p .

No other unbiased estimator of p has a smaller variance.

Example 2: Normal — MVUEs for μ and σ^2

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both μ and σ^2 unknown.

Example 2: Normal — MVUEs for μ and σ^2

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both μ and σ^2 unknown.

Step 1: Find the sufficient statistics.

The likelihood is:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right)$$

Example 2: Normal — MVUEs for μ and σ^2

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both μ and σ^2 unknown.

Step 1: Find the sufficient statistics.

The likelihood is:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right)$$

Expanding $(y_i - \mu)^2 = y_i^2 - 2\mu y_i + \mu^2$ and regrouping:

$$L = \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(\frac{-n\mu^2}{2\sigma^2} \right) \exp \left[-\frac{1}{2\sigma^2} \left(\sum y_i^2 - 2\mu \sum y_i \right) \right]}_{g(\sum y_i, \sum y_i^2, \mu, \sigma^2)} \cdot \underbrace{1}_{h(y)}$$

Example 2: Normal — MVUEs for μ and σ^2

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both μ and σ^2 unknown.

Step 1: Find the sufficient statistics.

The likelihood is:

$$L(\mu, \sigma^2) = \left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right)$$

Expanding $(y_i - \mu)^2 = y_i^2 - 2\mu y_i + \mu^2$ and regrouping:

$$L = \underbrace{\left(\frac{1}{\sigma\sqrt{2\pi}} \right)^n \exp \left(\frac{-n\mu^2}{2\sigma^2} \right) \exp \left[-\frac{1}{2\sigma^2} \left(\sum y_i^2 - 2\mu \sum y_i \right) \right]}_{g(\sum y_i, \sum y_i^2, \mu, \sigma^2)} \cdot \underbrace{1}_{h(y)}$$

By the factorization criterion: $\sum Y_i$ and $\sum Y_i^2$ are **jointly sufficient** for μ and σ^2 .
(Equivalently, $\bar{Y}$ and S^2 are jointly sufficient.)

Example 2 (continued): The MVUEs

Step 2: Find unbiased functions of the sufficient statistics.

Example 2 (continued): The MVUEs

Step 2: Find unbiased functions of the sufficient statistics.

For μ : We need an unbiased function of $\sum Y_i$. Since $\mathbb{E}[\bar{Y}] = \mu$:

$$\boxed{\hat{\mu}_{\text{MVUE}} = \bar{Y}} \quad \text{is the MVUE for } \mu$$

Example 2 (continued): The MVUEs

Step 2: Find unbiased functions of the sufficient statistics.

For μ : We need an unbiased function of $\sum Y_i$. Since $\mathbb{E}[\bar{Y}] = \mu$:

$$\boxed{\hat{\mu}_{\text{MVUE}} = \bar{Y}} \quad \text{is the MVUE for } \mu$$

For σ^2 : We need an unbiased function of $\sum Y_i$ and $\sum Y_i^2$. Recall:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n-1} \left[\sum Y_i^2 - n\bar{Y}^2 \right]$$

This is a function of $\sum Y_i$ and $\sum Y_i^2$, and $\mathbb{E}[S^2] = \sigma^2$:

$$\boxed{\hat{\sigma}_{\text{MVUE}}^2 = S^2} \quad \text{is the MVUE for } \sigma^2$$

Example 2 (continued): The MVUEs

Step 2: Find unbiased functions of the sufficient statistics.

For μ : We need an unbiased function of $\sum Y_i$. Since $\mathbb{E}[\bar{Y}] = \mu$:

$$\boxed{\hat{\mu}_{\text{MVUE}} = \bar{Y}} \quad \text{is the MVUE for } \mu$$

For σ^2 : We need an unbiased function of $\sum Y_i$ and $\sum Y_i^2$. Recall:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \bar{Y})^2 = \frac{1}{n-1} \left[\sum Y_i^2 - n\bar{Y}^2 \right]$$

This is a function of $\sum Y_i$ and $\sum Y_i^2$, and $\mathbb{E}[S^2] = \sigma^2$:

$$\boxed{\hat{\sigma}_{\text{MVUE}}^2 = S^2} \quad \text{is the MVUE for } \sigma^2$$

The familiar $\bar{Y}$ and S^2 are not just “good” estimators—they are the **best possible** unbiased estimators. No unbiased estimator of μ (or σ^2) can have smaller variance.

Example 3: Exponential — MVUE for a Function of θ

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\theta)$ with $f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$, $y > 0$.
Recall: $\mathbb{E}[Y_i] = \theta$ and $\text{Var}(Y_i) = \theta^2$.

Example 3: Exponential — MVUE for a Function of θ

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\theta)$ with $f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$, $y > 0$.

Recall: $\mathbb{E}[Y_i] = \theta$ and $\text{Var}(Y_i) = \theta^2$.

Question: Find the MVUE of $\text{Var}(Y_i) = \theta^2$.

Example 3: Exponential — MVUE for a Function of θ

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\theta)$ with $f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$, $y > 0$.

Recall: $\mathbb{E}[Y_i] = \theta$ and $\text{Var}(Y_i) = \theta^2$.

Question: Find the MVUE of $\text{Var}(Y_i) = \theta^2$.

Step 1: The sufficient statistic is $U = \sum Y_i$ (from the factorization criterion).

Since $\bar{Y} = U/n$ is a function of U , and $\mathbb{E}[\bar{Y}] = \theta$, we know $\bar{Y}$ is the MVUE for θ .

Example 3: Exponential — MVUE for a Function of θ

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\theta)$ with $f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$, $y > 0$.

Recall: $\mathbb{E}[Y_i] = \theta$ and $\text{Var}(Y_i) = \theta^2$.

Question: Find the MVUE of $\text{Var}(Y_i) = \theta^2$.

Step 1: The sufficient statistic is $U = \sum Y_i$ (from the factorization criterion).

Since $\bar{Y} = U/n$ is a function of U , and $\mathbb{E}[\bar{Y}] = \theta$, we know $\bar{Y}$ is the MVUE for θ .

Step 2: What about θ^2 ? A natural first guess is $\bar{Y}^2$. Is it unbiased?

$$\mathbb{E}[\bar{Y}^2] = \text{Var}(\bar{Y}) + [\mathbb{E}(\bar{Y})]^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n} \theta^2$$

Example 3: Exponential — MVUE for a Function of θ

Setup: $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Exponential}(\theta)$ with $f(y|\theta) = \frac{1}{\theta} e^{-y/\theta}$, $y > 0$.

Recall: $\mathbb{E}[Y_i] = \theta$ and $\text{Var}(Y_i) = \theta^2$.

Question: Find the MVUE of $\text{Var}(Y_i) = \theta^2$.

Step 1: The sufficient statistic is $U = \sum Y_i$ (from the factorization criterion).

Since $\bar{Y} = U/n$ is a function of U , and $\mathbb{E}[\bar{Y}] = \theta$, we know $\bar{Y}$ is the MVUE for θ .

Step 2: What about θ^2 ? A natural first guess is $\bar{Y}^2$. Is it unbiased?

$$\mathbb{E}[\bar{Y}^2] = \text{Var}(\bar{Y}) + [\mathbb{E}(\bar{Y})]^2 = \frac{\theta^2}{n} + \theta^2 = \frac{n+1}{n} \theta^2$$

$\bar{Y}^2$ is **biased**! It overestimates θ^2 by a factor of $\frac{n+1}{n}$.

Example 3 (continued): Correcting the Bias

Since $\mathbb{E}[\bar{Y}^2] = \frac{n+1}{n} \theta^2$, we can fix the bias by multiplying by $\frac{n}{n+1}$:

$$\mathbb{E}\left[\frac{n}{n+1} \bar{Y}^2\right] = \frac{n}{n+1} \cdot \frac{n+1}{n} \theta^2 = \theta^2 \quad \checkmark$$

Example 3 (continued): Correcting the Bias

Since $\mathbb{E}[\bar{Y}^2] = \frac{n+1}{n} \theta^2$, we can fix the bias by multiplying by $\frac{n}{n+1}$:

$$\mathbb{E}\left[\frac{n}{n+1} \bar{Y}^2\right] = \frac{n}{n+1} \cdot \frac{n+1}{n} \theta^2 = \theta^2 \quad \checkmark$$

This estimator is a function of $\bar{Y}$ (hence of $U = \sum Y_i$) $\checkmark$ and unbiased for θ^2 $\checkmark$

Example 3 (continued): Correcting the Bias

Since $\mathbb{E}[\bar{Y}^2] = \frac{n+1}{n} \theta^2$, we can fix the bias by multiplying by $\frac{n}{n+1}$:

$$\mathbb{E}\left[\frac{n}{n+1} \bar{Y}^2\right] = \frac{n}{n+1} \cdot \frac{n+1}{n} \theta^2 = \theta^2 \quad \checkmark$$

This estimator is a function of $\bar{Y}$ (hence of $U = \sum Y_i$) $\checkmark$ and unbiased for θ^2 $\checkmark$

Result

$$\hat{\theta}^2_{\text{MVUE}} = \frac{n}{n+1} \bar{Y}^2$$

is the MVUE of θ^2 . No unbiased estimator of θ^2 has a smaller variance.

Example 3 (continued): Correcting the Bias

Since $\mathbb{E}[\bar{Y}^2] = \frac{n+1}{n} \theta^2$, we can fix the bias by multiplying by $\frac{n}{n+1}$:

$$\mathbb{E}\left[\frac{n}{n+1} \bar{Y}^2\right] = \frac{n}{n+1} \cdot \frac{n+1}{n} \theta^2 = \theta^2 \quad \checkmark$$

This estimator is a function of $\bar{Y}$ (hence of $U = \sum Y_i$) $\checkmark$ and unbiased for θ^2 $\checkmark$

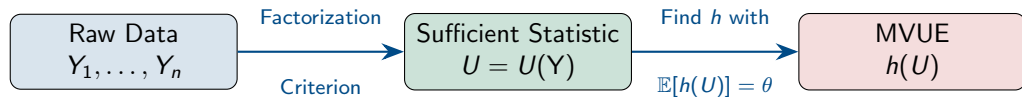
Result

$$\hat{\theta}^2_{\text{MVUE}} = \frac{n}{n+1} \bar{Y}^2$$

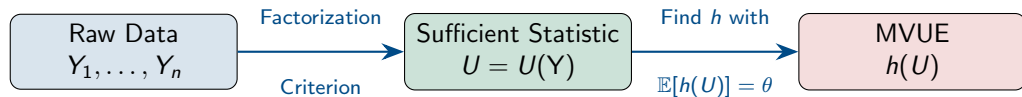
is the MVUE of θ^2 . No unbiased estimator of θ^2 has a smaller variance.

When estimating a **function** of a parameter (like θ^2), the plug-in estimator ($\bar{Y}^2$) may be biased. A simple correction—scaling by a constant—can fix it and still give the MVUE.

The Big Picture: From Data to the Best Estimator



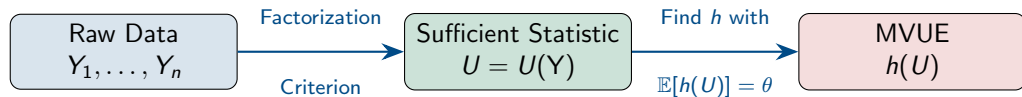
The Big Picture: From Data to the Best Estimator



The Estimation Pipeline

- 1 **Compress:** Use the factorization criterion to find U —captures *all* information about θ .
- 2 **Build:** Find $h(U)$ that is unbiased, i.e., $\mathbb{E}[h(U)] = \theta$.
- 3 **Done!** By Rao–Blackwell, $h(U)$ is the MVUE.

The Big Picture: From Data to the Best Estimator



The Estimation Pipeline

- 1 **Compress:** Use the factorization criterion to find U —captures *all* information about θ .
- 2 **Build:** Find $h(U)$ that is unbiased, i.e., $\mathbb{E}[h(U)] = \theta$.
- 3 **Done!** By Rao–Blackwell, $h(U)$ is the MVUE.

Think of it like data compression: the sufficient statistic compresses the data *without losing any information*. Building your estimator from this compressed data gives you the best possible precision.

Summary of MVUEs We Found

Distribution	Parameter	Sufficient Stat.	MVUE
Bernoulli(p)	p	$\sum Y_i$	$\bar{Y}$
$N(\mu, \sigma^2)$	μ	$\sum Y_i, \sum Y_i^2$	$\bar{Y}$
$N(\mu, \sigma^2)$	σ^2	$\sum Y_i, \sum Y_i^2$	S^2
Exponential(θ)	θ	$\sum Y_i$	$\bar{Y}$
Exponential(θ)	θ^2	$\sum Y_i$	$\frac{n}{n+1} \bar{Y}^2$

Summary of MVUEs We Found

Distribution	Parameter	Sufficient Stat.	MVUE
Bernoulli(p)	p	$\sum Y_i$	$\bar{Y}$
$N(\mu, \sigma^2)$	μ	$\sum Y_i, \sum Y_i^2$	$\bar{Y}$
$N(\mu, \sigma^2)$	σ^2	$\sum Y_i, \sum Y_i^2$	S^2
Exponential(θ)	θ	$\sum Y_i$	$\bar{Y}$
Exponential(θ)	θ^2	$\sum Y_i$	$\frac{n}{n+1} \bar{Y}^2$

Notice a pattern: $\bar{Y}$ appears everywhere! For many common distributions, the sample mean is the MVUE for the population mean. But for **functions** of parameters (like θ^2), you often need to adjust for bias.

Key Takeaways

- ① **Rao–Blackwell Theorem:** Conditioning any unbiased estimator on a sufficient statistic produces a new estimator that is still unbiased with variance $\leq$ the original.
- ② **MVUE = “the best unbiased estimator”:** It has the smallest variance among all unbiased estimators.
- ③ **Recipe:** Find the sufficient statistic via factorization, then find an unbiased function of it $\Rightarrow$ MVUE.
- ④ **Key examples:**
 - $\bar{Y}$ is the MVUE for μ (Normal) and for p (Bernoulli)
 - S^2 is the MVUE for σ^2 (Normal)
 - For functions of parameters, plug-in estimators may be biased—correct with a scaling constant

Coming Up Next Week

Week 5: Methods of Estimation—Maximum Likelihood

- Detailed treatment of Maximum Likelihood Estimation (MLE)
- Properties of MLEs: consistency, asymptotic normality, asymptotic efficiency
- Why MLEs are “asymptotically best”: they achieve maximum precision as $n \rightarrow \infty$
- Connection to this week: MLEs are functions of sufficient statistics!

So Far we covered

Week 1: CLT — sample averages are approximately normal

Week 2: Sampling distributions — the building blocks

Week 3: Point estimators

Week 4: *Today* — sufficient statistics, Rao–Blackwell, & MVUE

Week 5: *Next* — methods of estimation

Further Reading

- Wackerly, Mendenhall & Scheaffer, *Mathematical Statistics with Applications*, 7th ed.
 - Section 9.4: Sufficiency
 - Section 9.5: The Rao–Blackwell Theorem and MVUE
- Casella & Berger, *Statistical Inference*, 2nd ed.
 - Ch. 6.2: Sufficient Statistics and the Factorization Theorem
 - Ch. 7.3: The Rao–Blackwell Theorem
- Hogg, McKean & Craig, *Introduction to Mathematical Statistics*, 8th ed.
 - Ch. 7.2–7.3: Sufficiency and Rao–Blackwell