# Point Estimation

## Statistical Inference - Chapter 2

Prof. Abdelhakim Necir

Mohamed Khider University, Biskra
Faculty of Exact Sciences
Department of Mathematics

Academic Year 2026–2027

# What is Point Estimation?

## Objective

Learn about unknown characteristics (parameters) of a population from a random sample.

- Let $X_1, \ldots, X_n$ be a random sample from a distribution depending on an unknown parameter $\theta$.

## Distinction

**Point estimation** provides a single value.
**Confidence interval estimation** (Chapter 3) provides a range of values.

# What is Point Estimation?

## Objective

Learn about unknown characteristics (parameters) of a population from a random sample.

- Let $X_1, \ldots, X_n$ be a random sample from a distribution depending on an unknown parameter $\theta$.
- A **point estimator** is a statistic $T(X_1, \ldots, X_n)$ used to estimate $\theta$.

## Distinction

**Point estimation** provides a single value.
**Confidence interval estimation** (Chapter 3) provides a range of values.

# What is Point Estimation?

## Objective

Learn about unknown characteristics (parameters) of a population from a random sample.

- Let $X_1, \ldots, X_n$ be a random sample from a distribution depending on an unknown parameter $\theta$.
- A **point estimator** is a statistic $T(X_1, \ldots, X_n)$ used to estimate $\theta$.
- It assigns a **single numerical value** to the unknown parameter.

## Distinction

**Point estimation** provides a single value.
**Confidence interval estimation** (Chapter 3) provides a range of values.

# Estimator vs. Estimate & Frameworks

## Estimator vs. Estimate

- **Estimator** $T(X_1, \ldots, X_n)$: A random variable (before observing data).
- **Estimate** $t = T(x_1, \ldots, x_n)$: A fixed numerical value (after observing data).

## Parametric vs. Nonparametric

- **Parametric**: Distribution belongs to a finite-dimensional family (e.g., $\mathcal{N}(\mu, \sigma^2)$).
- **Nonparametric**: No finite-dimensional restriction (e.g., estimating a density).

# Maximum Likelihood Estimation (MLE)

## Principle

Choose the parameter value that makes the observed data most probable.

- **Likelihood function** (for independent observations):

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

# Maximum Likelihood Estimation (MLE)

## Principle

Choose the parameter value that makes the observed data most probable.

- **Likelihood function** (for independent observations):

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

- **Log-Likelihood** (easier to maximize):

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

# Maximum Likelihood Estimation (MLE)

## Principle

Choose the parameter value that makes the observed data most probable.

- **Likelihood function** (for independent observations):

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta)$$

- **Log-Likelihood** (easier to maximize):

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta)$$

- **MLE** $\widehat{\theta}_{\text{MLE}}$ is the value that maximizes $L(\theta)$ or $\ell(\theta)$:

$$\widehat{\theta}_{\text{MLE}} \in \arg\max_{\theta \in \Theta} \ell(\theta)$$

# Method of Moments (MoM)

## Principle

Equate sample moments to theoretical (population) moments.

- **Theoretical moment** of order $r$: $\mu'_r(\theta) = \mathbb{E}_\theta[X^r]$

# Method of Moments (MoM)

## Principle

Equate sample moments to theoretical (population) moments.

- **Theoretical moment** of order $r$: $\mu'_r(\theta) = \mathbb{E}_\theta[X^r]$
- **Sample moment** of order $r$: $m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$

# Method of Moments (MoM)

## Principle

Equate sample moments to theoretical (population) moments.

- **Theoretical moment** of order $r$: $\mu'_r(\theta) = \mathbb{E}_\theta[X^r]$
- **Sample moment** of order $r$: $m'_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
- **Method**: Solve the system of equations:

$$\mu'_r(\theta) = m'_r, \qquad r = 1, \ldots, k$$

where $k$ is the number of unknown parameters.

# Bias, Variance, and Mean Squared Error

## Bias

$$\text{Bias}_\theta(T) = \mathbb{E}_\theta[T] - \theta$$

$T$ is **unbiased** if $\mathbb{E}_\theta[T] = \theta$.

## Variance

$$\text{Var}_\theta(T) = \mathbb{E}_\theta\left[(T - \mathbb{E}_\theta[T])^2\right]$$

## Mean Squared Error (MSE)

$$\text{MSE}_\theta(T) = \mathbb{E}_\theta\left[(T-\theta)^2\right]$$

**Decomposition:**

$$\text{MSE} = \text{Var}(T) + [\text{Bias}(T)]^2$$

# Consistency and Efficiency

## Consistency (Asymptotic property)

An estimator $T_n$ is **consistent** for $\theta$ if it converges in probability to $\theta$ as $n \longrightarrow \infty$:

$$T_n \xrightarrow{\mathbb{P}} \theta \iff \lim_{n\to\infty} P_\theta\left(|T_n - \theta| > \varepsilon\right) = 0$$

## Efficiency

If $T_1$ and $T_2$ are unbiased estimators of $\theta$, $T_1$ is **more efficient** than $T_2$ if:

$$\mathrm{Var}_\theta(T_1) \leq \mathrm{Var}_\theta(T_2) \quad \forall \theta \in \Theta$$

# Sufficiency: Intuition and Definition

## Intuition

A statistic $T(X)$ is sufficient if it captures **all** the information in the sample relevant to $\theta$. Once $T(X)$ is known, the sample provides no further information about $\theta$.

## Formal Definition

$T(X)$ is **sufficient** for $\theta$ if the conditional distribution of $X$ given $T(X) = t$ is **independent of** $\theta$:

$$P_\theta(X = x \mid T(X) = t) \quad \text{does not depend on } \theta.$$

# Fisher-Neyman Factorization Theorem

## Theorem (Fisher-Neyman)

*A statistic $T(X)$ is sufficient for $\theta$ if and only if the joint density/mass function factors as:*

$$f(x_1, \ldots, x_n; \theta) = g\big(T(x), \theta\big)\, h(x_1, \ldots, x_n)$$

- $g(t, \theta) \geq 0$: depends on data **only** through $T(x)$ and on $\theta$.
- $h(x) \geq 0$: depends **only** on the data, **not** on $\theta$.

# Proof of Factorization Theorem ($\Rightarrow$): Setup

**Direct Implication:** Assume $f(x;\theta) = g(T(x),\theta)h(x)$. Prove $T$ is sufficient.

1. Express the event $\{T = t\}$:

$$\{T(X) = t\} = \bigcup_{x' : T(x') = t} \{X = x'\}$$

# Proof of Factorization Theorem ($\Rightarrow$): Setup

**Direct Implication:** Assume $f(x;\theta) = g(T(x),\theta)h(x)$. Prove $T$ is sufficient.

1. Express the event $\{T = t\}$:

$$\{T(X) = t\} = \bigcup_{x' : T(x') = t} \{X = x'\}$$

2. By $\sigma$-additivity, the marginal probability of $T = t$ is:

$$P_\theta(T = t) = \sum_{x' : T(x') = t} g(t,\theta)h(x') = g(t,\theta) \sum_{x' : T(x') = t} h(x')$$

# Proof of Factorization Theorem ($\Rightarrow$): Conclusion

3. The conditional probability is:

$$P_\theta(X = x \mid T = t) = \frac{g(t,\theta)h(x)}{g(t,\theta)\sum_{x'} h(x')} = \frac{h(x)}{\sum_{x'} h(x')}$$

**Conclusion:** The $g(t,\theta)$ terms cancel. The result is independent of $\theta$. ■

# Proof of Factorization Theorem ($\Leftarrow$): Setup

**Converse Implication:** Assume $T$ is sufficient. Prove the factorization holds.

1. By definition of sufficiency, the conditional probability is independent of $\theta$:

$$P_\theta(X = x \mid T = T(x)) = h(x)$$

# Proof of Factorization Theorem ($\Leftarrow$): Setup

**Converse Implication:** Assume $T$ is sufficient. Prove the factorization holds.

1. By definition of sufficiency, the conditional probability is independent of $\theta$:
$$P_\theta(X = x \mid T = T(x)) = h(x)$$
2. Using the definition of conditional probability:
$$P_\theta(X = x) = P_\theta(X = x \mid T = T(x)) \times P_\theta(T = T(x))$$

# Proof of Factorization Theorem ($\Leftarrow$): Conclusion

3. Define $g(t, \theta) := P_\theta(T = t)$. Then:

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

**Conclusion:** The joint density is factored. ■

# Examples: Bernoulli and Poisson

## Bernoulli: $X_i \sim \text{Bernoulli}(p)$

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

$\Rightarrow T = \sum X_i$ is sufficient for $p$.

## Poisson: $X_i \sim \text{Poisson}(\lambda)$

$$f(x;\lambda) = e^{-n\lambda}\lambda^{\sum x_i}\prod \frac{1}{x_i!} = \underbrace{e^{-n\lambda}\lambda^T}_{g(T,\lambda)} \cdot \underbrace{\prod \frac{1}{x_i!}}_{h(x)}$$

$\Rightarrow T = \sum X_i$ is sufficient for $\lambda$.

# Examples: Exponential and Normal

## Exponential: $X_i \sim \text{Exp}(\lambda)$

$$f(x;\lambda) = \lambda^n e^{-\lambda \sum x_i} \mathbf{1}_{\{x_i>0\}} = \underbrace{\lambda^n e^{-\lambda T}}_{g(T,\lambda)} \cdot \underbrace{\mathbf{1}_{\{x_i>0\}}}_{h(x)}$$

$\Rightarrow T = \sum X_i$ is sufficient for $\lambda$.

## Normal: $X_i \sim \mathcal{N}(\mu, \sigma^2)$ (both unknown)

Expanding $\sum (x_i - \mu)^2$, the joint density factors with:
$T(x) = \left(\sum x_i, \sum x_i^2\right)$.
$\Rightarrow T = \left(\sum X_i, \sum X_i^2\right)$ is sufficient for $(\mu, \sigma^2)$.

# Completeness: Definition and Intuition

## Definition

A statistic $T$ is **complete** for $\theta$ if for every measurable function $g$:

$$\mathbb{E}_\theta[g(T)] = 0 \ \ \forall \theta \in \Theta \quad \Longrightarrow \quad P_\theta(g(T) = 0) = 1 \ \ \forall \theta \in \Theta$$

## Interpretation

- The only function of $T$ with an expected value of zero for all $\theta$ is the zero function.
- **Why it matters:** If $T$ is sufficient and complete, any unbiased estimator based on $T$ is **unique**.

# Proof of Completeness: Bernoulli Model (1/3)

Let $X_i \sim \text{Bernoulli}(p)$. Then $T = \sum X_i \sim \text{Binomial}(n, p)$.
Assume $\mathbb{E}_p[g(T)] = 0$ for all $p \in (0, 1)$.

1. Write the expectation:

$$\sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t} = 0$$

# Proof of Completeness: Bernoulli Model (1/3)

Let $X_i \sim \text{Bernoulli}(p)$. Then $T = \sum X_i \sim \text{Binomial}(n, p)$.
Assume $\mathbb{E}_p[g(T)] = 0$ for all $p \in (0, 1)$.

1. Write the expectation:

$$\sum_{t=0}^{n} g(t) \binom{n}{t} p^t (1-p)^{n-t} = 0$$

2. Multiply by $\frac{1}{(1-p)^n}$:

$$\sum_{t=0}^{n} g(t) \binom{n}{t} \left(\frac{p}{1-p}\right)^t = 0$$

# Proof of Completeness: Bernoulli Model (2/3)

3. Let $u = \frac{p}{1-p}$. As $p \in (0,1)$, $u \in (0,+\infty)$. We get a polynomial in $u$:

$$\sum_{t=0}^{n} \left[ g(t) \binom{n}{t} \right] u^t = 0 \quad \forall u > 0$$

# Proof of Completeness: Bernoulli Model (3/3)

4. A polynomial that vanishes on an infinite set (here $u > 0$) must be identically zero. Therefore, all its coefficients are zero:

$$g(t)\binom{n}{t} = 0 \quad \text{for all } t \in \{0, 1, \ldots, n\}$$

$\Rightarrow T = \sum X_i$ is **complete** for $p$. ■

# Proof of Completeness: Bernoulli Model (3/3)

4. A polynomial that vanishes on an infinite set (here $u > 0$) must be identically zero. Therefore, all its coefficients are zero:

$$g(t)\binom{n}{t} = 0 \quad \text{for all } t \in \{0, 1, \ldots, n\}$$

5. Since $\binom{n}{t} > 0$, we conclude:

$$g(t) = 0 \quad \text{for all } t.$$

$\Rightarrow T = \sum X_i$ is **complete** for $p$. ■

# Proof of Completeness: Poisson Model (1/3)

Let $X_i \sim \text{Poisson}(\lambda)$. Then $T = \sum X_i \sim \text{Poisson}(n\lambda)$.
Assume $\mathbb{E}_\lambda[g(T)] = 0$ for all $\lambda > 0$.

1. Write the expectation:

$$\sum_{t=0}^{\infty} g(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} = 0$$

# Proof of Completeness: Poisson Model (1/3)

Let $X_i \sim \text{Poisson}(\lambda)$. Then $T = \sum X_i \sim \text{Poisson}(n\lambda)$.
Assume $\mathbb{E}_\lambda[g(T)] = 0$ for all $\lambda > 0$.

1. Write the expectation:

$$\sum_{t=0}^{\infty} g(t) \frac{e^{-n\lambda}(n\lambda)^t}{t!} = 0$$

2. Multiply by $e^{n\lambda}$:

$$\sum_{t=0}^{\infty} \left[ g(t) \frac{n^t}{t!} \right] \lambda^t = 0 \quad \forall \lambda > 0$$

# Proof of Completeness: Poisson Model (2/3)

3. This is a power series in $\lambda$ that vanishes on $(0, +\infty)$.

# Proof of Completeness: Poisson Model (3/3)

4. By the identity theorem for power series, if it vanishes on an interval, all coefficients must be zero:

$$g(t)\frac{n^t}{t!} = 0 \quad \text{for all } t \in \{0, 1, 2, \ldots\}$$

$\Rightarrow T = \sum X_i$ is **complete** for $\lambda$. ■

# Proof of Completeness: Poisson Model (3/3)

4. By the identity theorem for power series, if it vanishes on an interval, all coefficients must be zero:

$$g(t)\frac{n^t}{t!} = 0 \quad \text{for all } t \in \{0, 1, 2, \ldots\}$$

5. Since $\frac{n^t}{t!} > 0$, we conclude:

$$g(t) = 0 \quad \text{for all } t.$$

$\Rightarrow T = \sum X_i$ is **complete** for $\lambda$. ■

# Proof of Completeness: Exponential Model (1/3)

Let $X_i \sim \text{Exp}(\lambda)$. Then $T = \sum X_i \sim \text{Gamma}(n, \lambda)$.
Assume $\mathbb{E}_\lambda[g(T)] = 0$ for all $\lambda > 0$.

1. Write the expectation using the Gamma density:

$$\int_0^\infty g(t) \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt = 0$$

# Proof of Completeness: Exponential Model (1/3)

Let $X_i \sim \text{Exp}(\lambda)$. Then $T = \sum X_i \sim \text{Gamma}(n, \lambda)$.
Assume $\mathbb{E}_\lambda[g(T)] = 0$ for all $\lambda > 0$.

1. Write the expectation using the Gamma density:

$$\int_0^\infty g(t) \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} dt = 0$$

2. Multiply by $\frac{(n-1)!}{\lambda^n}$:

$$\int_0^\infty \left[ g(t) t^{n-1} \right] e^{-\lambda t} dt = 0 \quad \forall \lambda > 0$$

# Proof of Completeness: Exponential Model (2/3)

3. Recognize the **Laplace transform** of $h(t) = g(t)t^{n-1}$:

$$\mathcal{L}[h](\lambda) = 0 \quad \forall \lambda > 0$$

# Proof of Completeness: Exponential Model (3/3)

4. By the uniqueness theorem for Laplace transforms, if $\mathcal{L}[h](\lambda) = 0$ for all $\lambda > 0$, then $h(t) = 0$ almost everywhere.

$\Rightarrow T = \sum X_i$ is **complete** for $\lambda$. ■

# Proof of Completeness: Exponential Model (3/3)

4. By the uniqueness theorem for Laplace transforms, if $\mathcal{L}[h](\lambda) = 0$ for all $\lambda > 0$, then $h(t) = 0$ almost everywhere.
5. Therefore, $g(t)t^{n-1} = 0$ a.e. Since $t^{n-1} > 0$ for $t > 0$:

$$g(t) = 0 \quad \text{a.e. on } (0, +\infty).$$

$\Rightarrow T = \sum X_i$ is **complete** for $\lambda$. ■

# MVUE and Lehmann-Scheffé Theorem

## Minimum Variance Unbiased Estimator (MVUE)

An unbiased estimator $T$ is an MVUE if its variance is $\leq$ the variance of any other unbiased estimator, for all $\theta$.

## Theorem (Lehmann-Scheffé)

*If $T$ is a* **complete** *and* **sufficient** *statistic for $\theta$, and if $g(T)$ is an* **unbiased** *estimator of $\theta$, then $g(T)$ is the* **unique MVUE** *of $\theta$.*

# Proof of Lehmann-Scheffé (Step 1 & 2)

**Setup:** Let $\tilde{\theta}$ be any unbiased estimator. Define $\hat{\theta}(T) = \mathbb{E}_\theta[\tilde{\theta} \mid T]$.

1. **Unbiasedness (Rao-Blackwell):** By the tower property,
   $\mathbb{E}_\theta[\hat{\theta}(T)] = \mathbb{E}_\theta[\mathbb{E}_\theta[\tilde{\theta} \mid T]] = \mathbb{E}_\theta[\tilde{\theta}] = \theta$.
   Since $T$ is sufficient, $\hat{\theta}(T)$ depends only on $T$, not on $\theta$.

# Proof of Lehmann-Scheffé (Step 1 & 2)

**Setup:** Let $\tilde{\theta}$ be any unbiased estimator. Define $\hat{\theta}(T) = \mathbb{E}_\theta[\tilde{\theta} \mid T]$.

1. **Unbiasedness (Rao-Blackwell):** By the tower property,
   $\mathbb{E}_\theta[\hat{\theta}(T)] = \mathbb{E}_\theta[\mathbb{E}_\theta[\tilde{\theta} \mid T]] = \mathbb{E}_\theta[\tilde{\theta}] = \theta$.
   Since $T$ is sufficient, $\hat{\theta}(T)$ depends only on $T$, not on $\theta$.
2. **Variance Reduction:** By the law of total variance:

$$\mathrm{Var}_\theta(\tilde{\theta}) = \mathrm{Var}_\theta\left(\mathbb{E}[\tilde{\theta} \mid T]\right) + \mathbb{E}\left[\mathrm{Var}(\tilde{\theta} \mid T)\right]$$

$$\mathrm{Var}_\theta(\tilde{\theta}) = \mathrm{Var}_\theta(\hat{\theta}(T)) + \underbrace{\mathbb{E}\left[\mathrm{Var}(\tilde{\theta} \mid T)\right]}_{\geq 0}$$

$$\Rightarrow \mathrm{Var}_\theta(\hat{\theta}(T)) \leq \mathrm{Var}_\theta(\tilde{\theta}).$$

# Proof of Lehmann-Scheffé (Step 3: Uniqueness)

3. **Uniqueness via Completeness:**
   Suppose $\hat{\theta}_1(T)$ and $\hat{\theta}_2(T)$ are two unbiased estimators based on $T$. Define $\Delta(T) = \hat{\theta}_1(T) - \hat{\theta}_2(T)$. Since both are unbiased:

$$\mathbb{E}_\theta[\Delta(T)] = \theta - \theta = 0 \quad \forall \theta \in \Theta$$

Because $T$ is **complete**, this implies:

$$P_\theta(\Delta(T) = 0) = 1 \quad \forall \theta \in \Theta$$

**Conclusion:** $\hat{\theta}_1(T) = \hat{\theta}_2(T)$ almost surely. The MVUE is unique. ■

# Exponential Families: Structure

## Definition

A family of distributions $\{f(x;\theta) : \theta \in \Theta\}$ is a $k$-**parameter exponential family** if the density/pmf can be written as:

$$f(x;\theta) = h(x)\, c(\theta)\, \exp\left(\sum_{j=1}^{k} \eta_j(\theta)\, T_j(x)\right)$$

- $\eta_j(\theta)$: **natural parameters** (functions of $\theta$)
- $T_j(x)$: **natural sufficient statistics**
- $h(x) \geq 0$: depends only on $x$
- $c(\theta) > 0$: normalization constant

# Exponential Families: Key Theorem

## Theorem (Sufficiency and Completeness)

*Let $X_1, \ldots, X_n$ be an i.i.d. sample from a $k$-parameter exponential family. Then:*

1. **Sufficiency:** *The statistic $T = (T_1(X), \ldots, T_k(X))$ is* **sufficient** *for $\theta$.*
2. **Completeness:** *If the natural parameter space contains an* **open rectangle** *in $\mathbb{R}^k$ (the family is* **regular***), then $T$ is* **complete***.*

# Bernoulli: Likelihood Function

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bernoulli}(p)$ with $p \in (0, 1)$.

## Likelihood Function

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i}$$

# Bernoulli: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(p) = \exp\left(\sum x_i \log p + \left(n - \sum x_i\right) \log(1-p)\right)$$

# Bernoulli: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(p) = \exp\left(\sum x_i \log p + \left(n - \sum x_i\right) \log(1-p)\right)$$

2. **Expand and rearrange:**

$$L(p) = (1-p)^n \exp\left(\log\left(\frac{p}{1-p}\right) \sum_{i=1}^{n} x_i\right)$$

# Bernoulli: Canonical Form

$$L(p) = \underbrace{(1-p)^n}_{c(p)} \cdot \exp\left( \underbrace{\log\left(\frac{p}{1-p}\right)}_{\eta(p)} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{T(x)} \right)$$

## Identification

- **Natural parameter:** $\eta = \log\left(\frac{p}{1-p}\right)$ (logit function)
- **Natural sufficient statistic:** $T = \sum_{i=1}^{n} X_i$

**Natural parameter space:** As $p \in (0,1)$, $\eta \in \mathbb{R}$ (open set).
$\Rightarrow$ Bernoulli is **regular** $\Rightarrow$ $T = \sum X_i$ is complete and sufficient.

# Poisson: Likelihood Function

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Poisson}(\lambda)$ with $\lambda > 0$.

## Likelihood Function

$$L(\lambda) = \prod_{i=1}^{n} \frac{e^{-\lambda} \lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

# Poisson: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(\lambda) = \frac{1}{\prod x_i!} \exp\left( -n\lambda + \log(\lambda) \sum_{i=1}^{n} x_i \right)$$

# Poisson: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(\lambda) = \frac{1}{\prod x_i!} \exp\left( -n\lambda + \log(\lambda) \sum_{i=1}^{n} x_i \right)$$

2. **Rearrange:**

$$L(\lambda) = \underbrace{\frac{1}{\prod x_i!}}_{h(x)} \cdot \underbrace{e^{-n\lambda}}_{c(\lambda)} \cdot \exp\left( \underbrace{\log \lambda}_{\eta(\lambda)} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{T(x)} \right)$$

# Poisson: Canonical Form Identification

## Identification

- $h(x) = \dfrac{1}{\prod x_i!}$
- $c(\lambda) = e^{-n\lambda}$
- **Natural parameter:** $\eta = \log \lambda$
- **Natural sufficient statistic:** $T = \sum_{i=1}^{n} X_i$

**Natural parameter space:** As $\lambda \in (0, +\infty)$, $\eta \in \mathbb{R}$ (open set).
$\Rightarrow$ Poisson is **regular** $\Rightarrow$ $T = \sum X_i$ is complete and sufficient.

# Exponential: Likelihood Function

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$ with $\lambda > 0$.

## Likelihood Function

$$L(\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} \mathbf{1}_{\{x_i > 0\}} = \lambda^n e^{-\lambda \sum_{i=1}^{n} x_i} \mathbf{1}_{\{\min x_i > 0\}}$$

# Exponential: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(\lambda) = \mathbf{1}_{\{\min x_i > 0\}} \cdot \lambda^n \cdot \exp\left( -\lambda \sum_{i=1}^{n} x_i \right)$$

# Exponential: Rewriting as Exponential

1. **Rewrite using exponential:**

$$L(\lambda) = \mathbf{1}_{\{\min x_i > 0\}} \cdot \lambda^n \cdot \exp\left( -\lambda \sum_{i=1}^{n} x_i \right)$$

2. **Rearrange:**

$$L(\lambda) = \underbrace{\mathbf{1}_{\{\min x_i > 0\}}}_{h(x)} \cdot \underbrace{\lambda^n}_{c(\lambda)} \cdot \exp\left( \underbrace{(-\lambda)}_{\eta(\lambda)} \cdot \underbrace{\sum_{i=1}^{n} x_i}_{T(x)} \right)$$

# Exponential: Canonical Form Identification

## Identification

- $h(x) = \mathbf{1}_{\{\min x_i > 0\}}$
- $c(\lambda) = \lambda^n$
- **Natural parameter:** $\eta = -\lambda$
- **Natural sufficient statistic:** $T = \sum_{i=1}^{n} X_i$

**Natural parameter space:** As $\lambda \in (0, +\infty)$, $\eta \in (-\infty, 0)$ (open set).
$\Rightarrow$ Exponential is **regular** $\Rightarrow$ $T = \sum X_i$ is complete and sufficient.

# Normal: Likelihood Function

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\theta = (\mu, \sigma^2)$.

## Likelihood Function

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

# Normal: Expanding the Quadratic

1. **Expand the quadratic:**

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2$$

# Normal: Expanding the Quadratic

1. **Expand the quadratic:**

$$\sum_{i=1}^{n}(x_i-\mu)^2=\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2$$

2. **Substitute:**

$$L(\mu,\sigma^2)=(2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum x_i^2+\frac{\mu}{\sigma^2}\sum x_i-\frac{n\mu^2}{2\sigma^2}\right)$$

# Normal: Canonical Form

$$L(\mu,\sigma^2) = \underbrace{(2\pi)^{-n/2}}_{h(x)} \cdot \underbrace{(\sigma^2)^{-n/2}\exp\left(-\frac{n\mu^2}{2\sigma^2}\right)}_{c(\mu,\sigma^2)} \cdot \exp\left(\eta_1 \sum x_i + \eta_2 \sum x_i^2\right)$$

## Identification (2-parameter exponential family)

- **Natural parameters:** $\eta_1 = \dfrac{\mu}{\sigma^2}, \quad \eta_2 = -\dfrac{1}{2\sigma^2}$
- **Natural sufficient statistics:** $T_1 = \sum_{i=1}^{n} X_i, \quad T_2 = \sum_{i=1}^{n} X_i^2$

# Normal: Regularity Check

**Natural parameter space:** $\mathcal{H} = \{(\eta_1, \eta_2) \in \mathbb{R}^2 : \eta_2 < 0\}$ (open rectangle in $\mathbb{R}^2$).

$\Rightarrow$ Normal is **regular** $\Rightarrow$ $T = (\sum X_i, \sum X_i^2)$ is complete and sufficient.

# Summary: Usual Distributions in Exponential Families

| **Distribution** | $\eta(\theta)$ | $T(x)$ | **Regular?** |
|---|---|---|---|
| Bernoulli($p$) | $\log\left(\frac{p}{1-p}\right)$ | $\sum X_i$ | Yes ($\eta \in \mathbb{R}$) |
| Poisson($\lambda$) | $\log \lambda$ | $\sum X_i$ | Yes ($\eta \in \mathbb{R}$) |
| Exponential($\lambda$) | $-\lambda$ | $\sum X_i$ | Yes ($\eta < 0$) |
| Normal($\mu, \sigma^2$) | $\left(\frac{\mu}{\sigma^2}, -\frac{1}{2\sigma^2}\right)$ | $\left(\sum X_i, \sum X_i^2\right)$ | Yes ($\eta_2 < 0$) |

## Consequence

For all these distributions, the natural sufficient statistic $T$ (which is a sum) is **complete and sufficient**, so any unbiased estimator based on $T$ is the **unique MVUE** (by Lehmann-ScheffÃĺ).

# Usual Estimators in Practice

## Quantitative Variables

- **Mean** $\mu$: $\overline{X} = \frac{1}{n}\sum X_i$
- **Variance** $\sigma^2$: $S^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$
- **Std Dev**: $S = \sqrt{S^2}$

*Note: The $n-1$ denominator ensures $S^2$ is unbiased.*

## Qualitative Variables (Proportion)

Let $X_i \sim \text{Bernoulli}(p)$.

**Proportion** $p$:

$$\widehat{p} = \frac{1}{n}\sum X_i = \overline{X}$$

*This is also the MLE of $p$.*

# Properties of Usual Estimators

## Unbiasedness

- $\mathbb{E}[\overline{X}] = \mu \quad \Rightarrow \quad \overline{X}$ is unbiased for $\mu$.
- $\mathbb{E}[S^2] = \sigma^2 \quad \Rightarrow \quad S^2$ is unbiased for $\sigma^2$.
- $\mathbb{E}[\widehat{p}] = p \quad \Rightarrow \quad \widehat{p}$ is unbiased for $p$.

## Optimality in the Normal Model

For $X_i \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$:

- $\overline{X}$ is the **MVUE** of $\mu$ (function of the complete sufficient statistic).
- $S^2$ is the **MVUE** of $\sigma^2$.

# Transition: What We Have Established

## What we have established

The Lehmann-Scheffé theorem guarantees the **existence** and **uniqueness** of the MVUE (Minimum Variance Unbiased Estimator) when a complete sufficient statistic exists.

# Transition: The Fundamental Question

## The fundamental question

- We know the MVUE has the **smallest variance among all unbiased estimators**.
- But **what is this minimum variance exactly**?
- Is there an **absolute theoretical lower bound** that applies to *any* unbiased estimator, regardless of the construction method?

# Transition: The Answer

**Answer**

Yes: the **Cramér-Rao-Fréchet-Darmois bound**, built from the **score function** and **Fisher information**.

# Relationship: MVUE vs. Efficiency

## MVUE

- Unbiased estimator with **smallest variance among all unbiased estimators**.
- Obtained via **Lehmann-Scheffé** (complete + sufficient statistic).
- Variance may be $>$ CRLB.

## Efficient Estimator

- Unbiased estimator that **attains** the Cramér-Rao bound.
- Variance $= \dfrac{1}{I_n(\theta)}$ exactly.
- Always an MVUE.

## Key Relationship

$$\text{Efficient} \implies \text{MVUE} \qquad \text{but} \qquad \text{MVUE} \not\implies \text{Efficient}$$

An efficient estimator is always an MVUE, but an MVUE may have a variance strictly greater than the CRLB (the bound is not always attainable).

# The Score Function

## Definition

The **score function** $U(\theta; X)$ measures the local sensitivity of the log-likelihood:

$$U(\theta; X) = \frac{\partial}{\partial \theta} \log f(X; \theta)$$

For a sample: $U_n(\theta) = \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \log f(X_i; \theta)$.

## Key Property

Under regularity conditions, the score has **zero expectation**:

$$\mathbb{E}_\theta[U(\theta; X)] = 0$$

**Consequence:** The MLE satisfies the first-order condition $U_n(\widehat{\theta}_{\text{MLE}}) = 0$.

# Proof: $\mathbb{E}_\theta[U(\theta; X)] = 0$ (1/2)

**Regularity conditions:** The support of $f(x;\theta)$ does not depend on $\theta$, and differentiation under the integral sign is valid.

1. Since $f(x;\theta)$ is a density:

$$\int f(x;\theta)\, dx = 1$$

# Proof: $\mathbb{E}_\theta[U(\theta; X)] = 0$ (1/2)

**Regularity conditions:** The support of $f(x;\theta)$ does not depend on $\theta$, and differentiation under the integral sign is valid.

1. Since $f(x;\theta)$ is a density:

$$\int f(x;\theta)\, dx = 1$$

2. Differentiate both sides with respect to $\theta$:

$$\frac{\partial}{\partial\theta} \int f(x;\theta)\, dx = 0$$

# Proof: $\mathbb{E}_\theta[U(\theta; X)] = 0$ (2/2)

3. Exchange derivative and integral (regularity):

$$\int \frac{\partial}{\partial \theta} f(x; \theta)\, dx = 0$$

# Proof: $\mathbb{E}_\theta[U(\theta; X)] = 0$ (2/2)

3. Exchange derivative and integral (regularity):

$$\int \frac{\partial}{\partial \theta} f(x; \theta)\, dx = 0$$

4. Rewrite using $\frac{\partial f}{\partial \theta} = f \cdot \frac{\partial \log f}{\partial \theta}$:

$$\int \frac{\partial \log f(x; \theta)}{\partial \theta} \cdot f(x; \theta)\, dx = 0$$

# Proof: $\mathbb{E}_\theta[U(\theta; X)] = 0$ (2/2)

3. Exchange derivative and integral (regularity):

$$\int \frac{\partial}{\partial \theta} f(x; \theta)\, dx = 0$$

4. Rewrite using $\dfrac{\partial f}{\partial \theta} = f \cdot \dfrac{\partial \log f}{\partial \theta}$:

$$\int \frac{\partial \log f(x; \theta)}{\partial \theta} \cdot f(x; \theta)\, dx = 0$$

5. Recognize the expectation:

$$\mathbb{E}_\theta[U(\theta; X)] = 0 \qquad \blacksquare$$

# Fisher Information

## Definition

The **Fisher Information** $I(\theta)$ quantifies the amount of information about $\theta$ in one observation:

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

For $n$ i.i.d. observations: $I_n(\theta) = n\, I(\theta)$.

## Interpretation

- $I(\theta) = \mathrm{Var}_\theta(U(\theta; X))$ since $\mathbb{E}[U] = 0$.
- High $I(\theta) \Rightarrow$ data is very informative about $\theta \Rightarrow$ precise estimation.
- $I_n(\theta)$ grows linearly with $n$: more data = more information.

# Example: Fisher Information for Bernoulli (1/2)

Let $X \sim \text{Bernoulli}(p)$. The pmf is $f(x;p) = p^x(1-p)^{1-x}$ for $x \in \{0,1\}$.

1. **Log-likelihood (one observation):**

$$\log f(x;p) = x \log p + (1-x)\log(1-p)$$

# Example: Fisher Information for Bernoulli (1/2)

Let $X \sim \text{Bernoulli}(p)$. The pmf is $f(x;p) = p^x(1-p)^{1-x}$ for $x \in \{0,1\}$.

1. **Log-likelihood (one observation):**

$$\log f(x;p) = x \log p + (1-x)\log(1-p)$$

2. **Score function:**

$$U(p;X) = \frac{\partial}{\partial p} \log f(X;p) = \frac{X}{p} - \frac{1-X}{1-p} = \frac{X-p}{p(1-p)}$$

# Example: Fisher Information for Bernoulli (2/2)

3. **Fisher information:**

$$I(p) = \mathrm{Var}_p(U) = \frac{\mathrm{Var}_p(X)}{[p(1-p)]^2} = \frac{p(1-p)}{p^2(1-p)^2} = \frac{1}{p(1-p)}$$

For $n$ observations: $I_n(p) = \dfrac{n}{p(1-p)}$.

# Example: Fisher Information for Exponential (1/2)

Let $X \sim \text{Exp}(\lambda)$. The density is $f(x;\lambda) = \lambda e^{-\lambda x}$ for $x > 0$.

1. **Log-density:**

$$\log f(x;\lambda) = \log \lambda - \lambda x$$

# Example: Fisher Information for Exponential (1/2)

Let $X \sim \text{Exp}(\lambda)$. The density is $f(x;\lambda) = \lambda e^{-\lambda x}$ for $x > 0$.

1. **Log-density:**
$$\log f(x;\lambda) = \log \lambda - \lambda x$$
2. **Score function:**
$$U(\lambda; X) = \frac{\partial}{\partial \lambda} \log f(X;\lambda) = \frac{1}{\lambda} - X$$

# Example: Fisher Information for Exponential (2/2)

3. **Fisher information:**

$$I(\lambda) = \mathrm{Var}_\lambda\left(\frac{1}{\lambda} - X\right) = \mathrm{Var}_\lambda(X) = \frac{1}{\lambda^2}$$

For $n$ observations: $I_n(\lambda) = \dfrac{n}{\lambda^2}$.

# The Cramér-Rao Lower Bound (CRLB)

## Theorem (Cramér-Rao Bound)

*For a regular model with Fisher information $I_n(\theta)$, if $T$ is an* **unbiased** *estimator of $\theta$:*

$$\mathrm{Var}_\theta(T) \geq \frac{1}{I_n(\theta)} = \frac{1}{n\,I(\theta)}$$

## Efficiency

- The bound provides a theoretical benchmark for precision.
- If an unbiased estimator **attains** the bound, it is called **efficient**.

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (1/2)

Let $T$ be an unbiased estimator ($\mathbb{E}_\theta[T] = \theta$) and $U_n$ the score ($\mathbb{E}_\theta[U_n] = 0$).

1. **Definition of covariance:**

$$\mathrm{Cov}_\theta(T, U_n) = \mathbb{E}_\theta[T \cdot U_n] - \mathbb{E}_\theta[T]\mathbb{E}_\theta[U_n] = \mathbb{E}_\theta[T \cdot U_n]$$

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (1/2)

Let $T$ be an unbiased estimator ($\mathbb{E}_\theta[T] = \theta$) and $U_n$ the score ($\mathbb{E}_\theta[U_n] = 0$).

1. **Definition of covariance:**

$$\mathrm{Cov}_\theta(T, U_n) = \mathbb{E}_\theta[T \cdot U_n] - \mathbb{E}_\theta[T]\mathbb{E}_\theta[U_n] = \mathbb{E}_\theta[T \cdot U_n]$$

2. **Write as an integral** using the likelihood $L(\theta; \mathbf{x})$:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \cdot U_n(\theta) \cdot L(\theta; \mathbf{x})\, d\mathbf{x}$$

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (1/2)

Let $T$ be an unbiased estimator ($\mathbb{E}_\theta[T] = \theta$) and $U_n$ the score ($\mathbb{E}_\theta[U_n] = 0$).

1. **Definition of covariance:**

$$\mathrm{Cov}_\theta(T, U_n) = \mathbb{E}_\theta[T \cdot U_n] - \mathbb{E}_\theta[T]\mathbb{E}_\theta[U_n] = \mathbb{E}_\theta[T \cdot U_n]$$

2. **Write as an integral** using the likelihood $L(\theta; \mathbf{x})$:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \cdot U_n(\theta) \cdot L(\theta; \mathbf{x})\, d\mathbf{x}$$

3. **Substitute the score** $U_n = \frac{1}{L}\frac{\partial L}{\partial \theta}$:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta}\, d\mathbf{x}$$

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (2/2)

Continuing from the previous slide:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x}$$

4. **Leibniz Rule (Regularity conditions):** Swap the derivative and the integral:

$$\int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x} = \frac{\partial}{\partial \theta} \left( \int T(\mathbf{x}) L(\theta; \mathbf{x}) \, d\mathbf{x} \right)$$

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (2/2)

Continuing from the previous slide:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x}$$

4. **Leibniz Rule (Regularity conditions):** Swap the derivative and the integral:

$$\int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x} = \frac{\partial}{\partial \theta} \left( \int T(\mathbf{x}) L(\theta; \mathbf{x}) \, d\mathbf{x} \right)$$

5. **Recognize the expectation:** The integral inside is exactly $\mathbb{E}_\theta[T]$.

$$\mathrm{Cov}_\theta(T, U_n) = \frac{\partial}{\partial \theta} \left( \mathbb{E}_\theta[T] \right)$$

# Proof: $\mathrm{Cov}_\theta(T, U_n) = 1$ (2/2)

Continuing from the previous slide:

$$\mathbb{E}_\theta[T \cdot U_n] = \int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x}$$

4. **Leibniz Rule (Regularity conditions):** Swap the derivative and the integral:

$$\int T(\mathbf{x}) \frac{\partial L(\theta; \mathbf{x})}{\partial \theta} \, d\mathbf{x} = \frac{\partial}{\partial \theta} \left( \int T(\mathbf{x}) L(\theta; \mathbf{x}) \, d\mathbf{x} \right)$$

5. **Recognize the expectation:** The integral inside is exactly $\mathbb{E}_\theta[T]$.

$$\mathrm{Cov}_\theta(T, U_n) = \frac{\partial}{\partial \theta} \left( \mathbb{E}_\theta[T] \right)$$

6. **Apply unbiasedness:** Since $\mathbb{E}_\theta[T] = \theta$, we get:

$$\mathrm{Cov}_\theta(T, U_n) = \frac{\partial}{\partial \theta}(\theta) = 1 \qquad \blacksquare$$

# Proof of the Cramér-Rao Bound (Cauchy-Schwarz)

1. **Apply Cauchy-Schwarz inequality:**

$$[\mathrm{Cov}_\theta(T, U_n)]^2 \leq \mathrm{Var}_\theta(T) \cdot \mathrm{Var}_\theta(U_n)$$

# Proof of the Cramér-Rao Bound (Cauchy-Schwarz)

1. **Apply Cauchy-Schwarz inequality:**

$$[\mathrm{Cov}_\theta(T, U_n)]^2 \leq \mathrm{Var}_\theta(T) \cdot \mathrm{Var}_\theta(U_n)$$

2. Since $\mathrm{Cov}_\theta(T, U_n) = 1$ and $\mathrm{Var}_\theta(U_n) = I_n(\theta)$:

$$1 \leq \mathrm{Var}_\theta(T) \cdot I_n(\theta) \quad \Rightarrow \quad \mathrm{Var}_\theta(T) \geq \frac{1}{I_n(\theta)} \qquad \blacksquare$$

# Relative Efficiency

## Definition

The **efficiency** of an unbiased estimator $T$ is defined as:

$$e(T) = \frac{1/I_n(\theta)}{\mathrm{Var}_\theta(T)} \in (0, 1]$$

- $e(T) = 1 \quad \Leftrightarrow \quad T$ is **efficient** (attains the CRLB).
- $e(T)$ close to $1 \Rightarrow T$ is nearly optimal.
- $e(T)$ small $\Rightarrow T$ is far from optimal.

# Relative Efficiency: Comparison

## Comparison of two unbiased estimators

The **relative efficiency** of $T_1$ with respect to $T_2$ is:

$$e(T_1, T_2) = \frac{\text{Var}_\theta(T_2)}{\text{Var}_\theta(T_1)}$$

$T_1$ is preferred if $e(T_1, T_2) > 1$.

# Example: Normal Mean (1/2)

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\sigma^2$ known.

- **Score:** $U(\mu; X) = \frac{X - \mu}{\sigma^2}$

# Example: Normal Mean (1/2)

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\sigma^2$ known.

- **Score:** $U(\mu; X) = \frac{X-\mu}{\sigma^2}$
- **Fisher Info (1 obs):** $I(\mu) = \mathbb{E}\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}$

# Example: Normal Mean (1/2)

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ with $\sigma^2$ known.

- **Score:** $U(\mu; X) = \frac{X-\mu}{\sigma^2}$
- **Fisher Info (1 obs):** $I(\mu) = \mathbb{E}\left[\frac{(X-\mu)^2}{\sigma^4}\right] = \frac{1}{\sigma^2}$
- **Fisher Info (n obs):** $I_n(\mu) = \frac{n}{\sigma^2}$

# Example: Normal Mean (2/2)

- **CRLB:** $\text{Var}_\mu(T) \geq \frac{\sigma^2}{n}$

## Conclusion

Since $\text{Var}(\overline{X}) = \frac{\sigma^2}{n}$, the sample mean **attains the bound**:

$$e(\overline{X}) = \frac{\sigma^2/n}{\sigma^2/n} = 1$$

$\overline{X}$ is **efficient** and therefore the MVUE of $\mu$.

# Summary of Chapter 2

## Construction of Estimators

- **MLE**: maximizes the likelihood (asymptotically efficient).
- **MoM**: equates sample and theoretical moments.

## Optimality Criteria

- **Sufficiency + Completeness** $\Rightarrow$ unique MVUE (Lehmann-ScheffÃĺ).
- **Fisher Information** $\Rightarrow$ CramÃĺr-Rao bound.
- **Efficiency**: unbiased + attains CRLB.

# Summary: Hierarchy of Optimality

## Hierarchy of Optimality

$$\text{Efficient} \implies \text{MVUE} \implies \text{Best among unbiased} \implies \text{Consistent}$$

# Key Takeaways (1/2)

1. **Sufficiency** reduces data without losing information about $\theta$.
2. **Completeness** guarantees uniqueness of unbiased estimators.
3. **Exponential families** provide a unified framework: regular $\Rightarrow$ complete + sufficient.

# Key Takeaways (2/2)

1.
4. **Lehmann-Scheffé** constructs the MVUE from a complete sufficient statistic.
5. **Fisher information** quantifies the theoretical precision limit.
6. **Cramér-Rao bound** provides the absolute lower bound on variance.

*These concepts form the foundation for Chapter 3:*
**Confidence Intervals and Hypothesis Testing.**