MODULE 4

# Estimation of Parameters

*...an unbiased estimator clusters tightly around the true value.*

# What We'll Cover

From a single best guess to a full confidence interval

1. **Point & Interval Estimation**
   A single number vs. a range with a confidence level

2. **Properties of a Good Estimator**
   Unbiasedness, minimum variance, consistency

3. **Maximum Likelihood Estimation**
   The method behind most estimators you'll ever use

4. **MLE for Binomial, Poisson, Normal**
   Deriving real estimators from real likelihoods

5. **Confidence Limits for the Mean**
   Turning a standard error into a usable interval

6. **Confidence Limits for a Proportion**
   The same idea, for categorical data

# Point Estimation

*A single number, computed from a sample, used as our best guess for a parameter*

Point estimate $\theta$ = a single value calculated from sample data, used to estimate a population parameter $\theta$

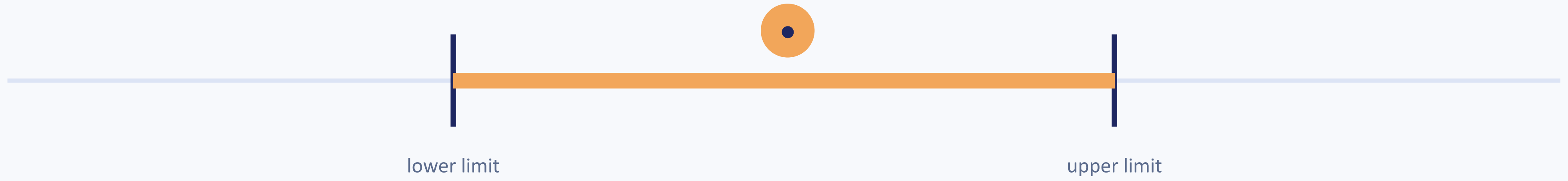
| Parameter $\theta$ | Common Point Estimator $\theta$ |
|---|---|
| Population mean $\mu$ | Sample mean $\bar{x}$ |
| Population variance $\sigma^2$ | Sample variance $s^2$ |
| Population proportion P | Sample proportion $\hat{p}$ |

**The catch: a point estimate is almost certainly not exactly right — it gives no sense of how far off it might be. That's exactly the gap interval estimation fills.**

# Interval Estimation

**A range of plausible values, together with a confidence level**

$$\text{Confidence interval} = \hat{\theta} \pm (\text{critical value}) \times SE(\hat{\theta})$$




**Correct interpretation:** a 95% confidence interval means that if we repeated this sampling process many times, about 95% of the intervals we'd construct would contain the true parameter — it is NOT a 95% probability that this particular interval contains θ.

**Width depends on:** the confidence level (higher confidence → wider interval), the standard error (more variability → wider interval), and the sample size (larger n → narrower interval).

# What Makes a Good Estimator?

*Three properties statisticians look for before trusting an estimator*

1

## Unbiased

On average, across all possible samples, it hits the true parameter value — $E(\theta) = \theta$

2

## Minimum Variance

Among unbiased estimators, it has the smallest possible spread — the most efficient choice

3

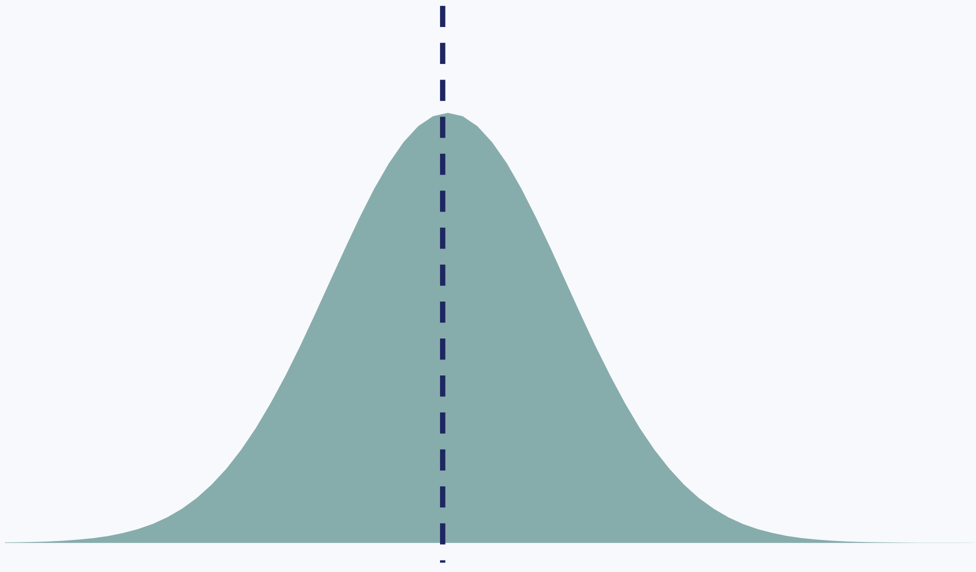
## Consistent

As the sample size grows, it gets closer and closer to the true parameter value

# Biased vs. Unbiased Estimators

*Where does the sampling distribution of $\vartheta$ sit, relative to the true $\vartheta$?*

### Unbiased Estimator




*Peak sits exactly on the true value: $E(\vartheta) = \vartheta$*

### Biased Estimator

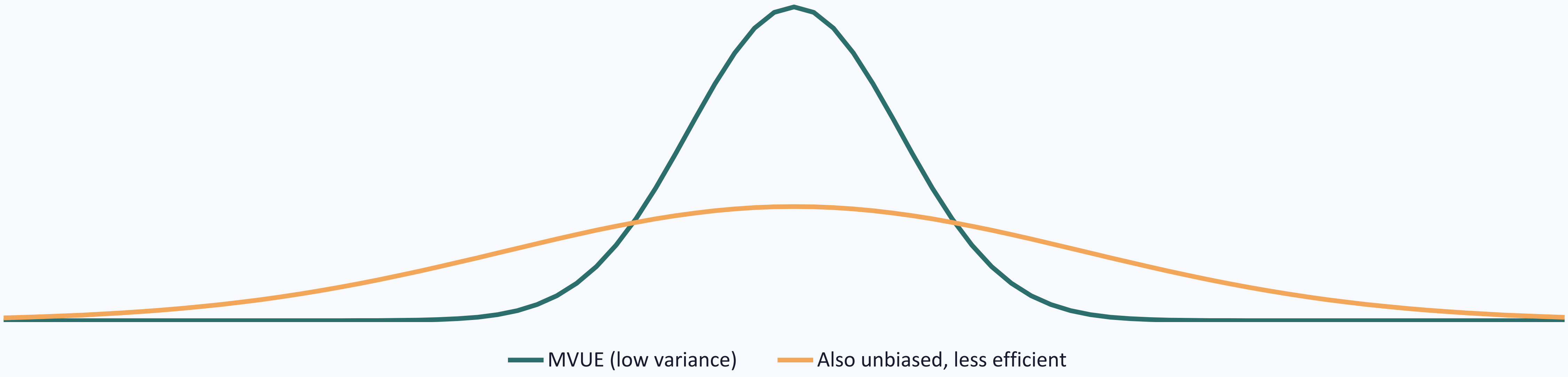



*Peak sits away from the true value: $E(\vartheta) \neq \vartheta$*

**Example: $s^2 = \Sigma(xi - \bar{x})^2/(n-1)$ is unbiased for $\sigma^2$, but dividing by n instead would be biased — it systematically underestimates $\sigma^2$ (see Module 3).**

# Minimum Variance Unbiased Estimator (MVUE)

*Among all unbiased estimators, the one with the smallest spread*




**Both are unbiased:** both curves are centred on the true θ — but the green curve is tighter, meaning it's more likely to land close to θ on any given sample. It is the MVUE.

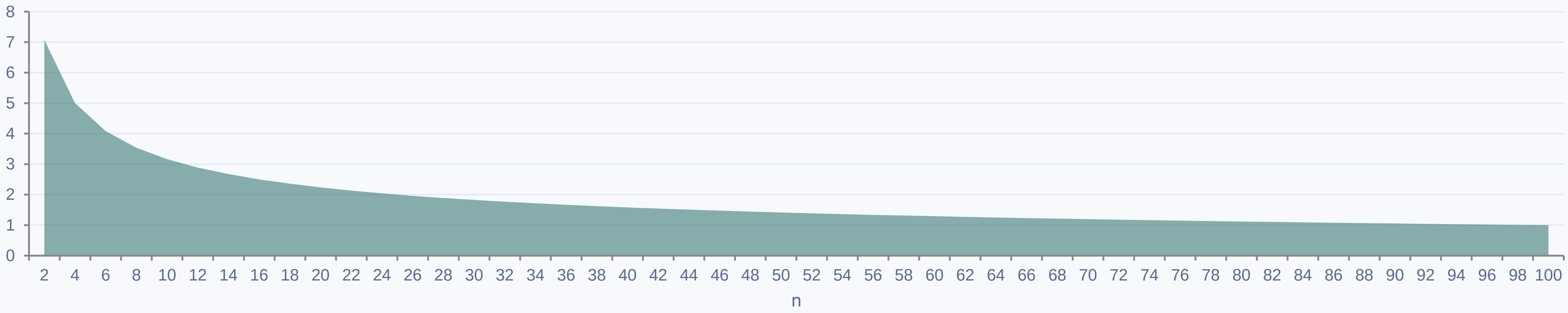
**Why it matters: given a choice between two unbiased estimators, always prefer the one with lower variance.**

# Consistent Estimator

*As the sample grows, the estimator homes in on the true parameter*

$\hat{\theta}_n \to \theta$ as $n \to \infty$ (both the bias and the variance shrink to zero)

**SE(x̄) = σ/√n as sample size grows (σ = 10)**




**Note: consistency and unbiasedness are different properties — an estimator can be biased for small n but still consistent, as long as the bias vanishes as n grows.**

# Maximum Likelihood Estimation — The Big Idea

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

1. You observe some data — e.g. 17 out of 20 IoT devices passed a test

2. For each possible value of the unknown parameter (e.g. $p$), ask: “how likely was this exact data, if $p$ had that value?”

3. The value of $p$ that makes the observed data most likely is the maximum likelihood estimate, $\hat{p}$

**Why it's so widely used: MLE is a single, general recipe that works across almost any distribution — Binomial, Poisson, Normal, and many more — and the resulting estimators are usually consistent and efficient for large samples.**

# The Method of Maximum Likelihood — Step by Step

*A recipe you can apply to almost any distribution*

1. 1) Write the likelihood: $L(\theta) = f(x_1;\theta)\cdot f(x_2;\theta)\cdot\cdots\cdot f(x_n;\theta)$ (product, since observations are independent)
2. 2) Take logs: $\ell(\theta) = \ln L(\theta)$ (sums are far easier to differentiate than products)
3. 3) Differentiate: $d\ell(\theta)/d\theta$
4. 4) Set to zero and solve: $d\ell(\theta)/d\theta = 0 \rightarrow \theta$ (the maximum likelihood estimate)

**Why take logs first? ln turns a product of many terms into a sum, which is far simpler to differentiate — and because ln is an increasing function, the θ that maximizes ln L(θ) is exactly the same θ that maximizes L(θ) itself.**

# MLE for the Binomial Parameter p

DERIVATION

*Deriving the estimator you've been using intuitively all along*

**Setup**

1. Likelihood: $(p) = C(n,x)\cdot p^{x}\cdot(1-p)^{n-x}$
2. Log-likelihood: $\ell(p) = \ln C(n,x) + x\cdot\ln p + (n-x)\cdot\ln(1-p)$
3. Differentiate: $d\ell/dp = x/p - (n-x)/(1-p)$
4. Set to zero: $(1-p) = (n-x)p \;\rightarrow\; x = np$

$$\hat{p} = x / n$$

**Worked check: 17 of 20 tested IoT devices pass quality control. The MLE for the pass probability is simply $\hat{p} = 17/20 = 0.85$ — exactly the sample proportion you'd have guessed intuitively, now formally justified.**

# MLE for the Poisson Parameter λ

DERIVATION

*The rate parameter's maximum likelihood estimate is the sample mean*

**Setup**

1. Likelihood: $(\lambda) = \Pi\ e^{-}\lambda\lambda^{xi}/xi!$
2. Log-likelihood: $\ell(\lambda) = -n\lambda + (\Sigma xi)\cdot\ln\lambda - \Sigma\ln(xi!)$
3. Differentiate: $d\ell/d\lambda = -n + \Sigma xi/\lambda$
4. Set to zero: $n = \Sigma xi/\lambda \rightarrow \hat{\lambda} = \Sigma xi/n = \bar{x}$

**$\hat{\lambda} = \bar{x}$ (the sample mean)**

**Worked check:** Intrusion attempts logged over 8 hours: 3, 5, 2, 4, 6, 3, 5, 4 (sum = 32). $\hat{\lambda} = 32/8 = 4$ attempts/hour — the same estimate we used for the Poisson worked examples in Module 1.

# MLE for the Normal Parameters μ and $\sigma^2$

DERIVATION

*Two parameters at once — and a familiar callback to Module 3*

Maximizing the Normal log-likelihood with respect to both parameters gives:

$$\hat{\mu} = \bar{x} \qquad \hat{\sigma}^2 = \Sigma(x_i - \bar{x})^2 / n$$

**A notable catch:** the MLE for the variance divides by n, not n−1 — which makes $\hat{\sigma}^2_{mle}$ a biased estimator, even though it is the maximum likelihood choice! This is exactly why Module 3 introduced $s^2$ with divisor n−1 as the preferred unbiased estimator.

**Worked check:** Sensor readings: 48, 52, 50, 49, 51 (mean = 50). $\hat{\sigma}^2 = [(48-50)^2+(52-50)^2+(50-50)^2+(49-50)^2+(51-50)^2]/5 = [4+4+0+1+1]/5 = 2$, so $\hat{\sigma} \approx 1.414$.

# MLE Estimators at a Glance

*Three distributions, three maximum likelihood estimators*

| Distribution | Parameter | MLE | In Words |
|---|---|---|---|
| **Binomial(n, p)** | **p** | $\hat{p} = x/n$ | sample proportion of successes |
| **Poisson(λ)** | **λ** | $\hat{\lambda} = \bar{x}$ | sample mean |
| **Normal(μ, σ²)** | **μ, σ²** | $\hat{\mu} = \bar{x},\ \hat{\sigma}^2 = \Sigma(xi - \bar{x})^2/n$ | sample mean; variance divided by n (biased) |

**The pattern across all three:** in every case, MLE for a location-type parameter (p, λ, or μ) turns out to be some form of sample average — which is exactly why the sample mean shows up everywhere in statistics: it is very often the maximum likelihood choice.

# Critical Values for Common Confidence Levels

*The z-values that appear in almost every confidence interval you'll compute*

| Confidence Level | α (uncovered) | Critical Value z |
|---|---|---|
| 90% | 0.10 | **1.645** |
| 95% | 0.05 | **1.960** |
| 99% | 0.01 | **2.576** |

**Where these come from:** each z-value marks off the middle portion of the standard Normal curve — e.g. exactly 95% of a Normal distribution's area lies between $z = -1.96$ and $z = +1.96$.

**Practical default:** 95% confidence ($z = 1.96$) is the overwhelming default in engineering, business, and scientific reporting unless a field specifically calls for something stricter or looser.

# Confidence Limits for the Population Mean

*Turning a point estimate and a standard error into a usable range*

$$\bar{x} - z\cdot(\sigma/\sqrt{n}) \le \mu \le \bar{x} + z\cdot(\sigma/\sqrt{n})$$

$\bar{x}$

**the sample mean — our point estimate of $\mu$**

$z$

**the critical value for the chosen confidence level (e.g. 1.96 for 95%)**

$\sigma/\sqrt{n}$

**the standard error of the mean, from Module 3**

If $\sigma$ is unknown

**substitute the sample standard deviation $s$ — valid for reasonably large $n$**

# Worked Example — Confidence Interval for a Mean

EXAMPLE

*Estimating average battery life for a fleet of IoT sensors*

**Question:** A sample of $n = 64$ sensors has mean battery life $\bar{x} = 100$ days, with population $\sigma = 16$ days. Construct a 95% confidence interval for $\mu$.

## Step-by-Step

1. **Standard error:** $SE = 16/\sqrt{64} = 2.00$
2. **Margin of error:** $1.96 \times 2 = 3.92$
3. **Interval:** $100 \pm 3.92$

**95% CI: (96.08, 103.92) days**

**Interpretation:** we are 95% confident the true average battery life across the whole fleet lies between 96.08 and 103.92 days — meaning about 95% of intervals built this way, from repeated sampling, would capture the true fleet average.

# Confidence Limits for the Population Proportion

*Same logic as the mean, adapted for a 0/1 outcome*

$$\hat{p} - z\cdot\sqrt{(\hat{p}\hat{q}/n)} \;\le\; P \;\le\; \hat{p} + z\cdot\sqrt{(\hat{p}\hat{q}/n)}$$

$\hat{p}$

**the sample proportion — our point estimate of P**

$\hat{q}$

**$1 - \hat{p}$**

z

**the same critical values as before (1.645 / 1.96 / 2.576)**

Sample size rule

**the Normal approximation needs $n\cdot\hat{p} \ge 5$ and $n\cdot\hat{q} \ge 5$**

# Worked Example — Confidence Interval for a Proportion

EXAMPLE

*Estimating the true failure rate from a sample audit*

**Question:** Out of $n = 400$ audited blockchain transactions, 88 failed validation ($\hat{p} = 0.22$). Construct a 95% confidence interval for the true failure rate P.

### Step-by-Step

1. Standard error: $SE = \sqrt{0.22 \times 0.78 / 400} = 0.0207$
2. Margin of error: $1.96 \times 0.0207 = 0.0406$
3. Interval: $0.22 \pm 0.0406$

**95% CI: (0.1794, 0.2606)**

**Interpretation:** we are 95% confident the true validation-failure rate across all transactions lies between about 17.9% and 26.1% — a fairly wide range, reflecting the moderate sample size of 400.

# Common Pitfalls to Avoid

*Mistakes that show up again and again with estimation*

### Misreading confidence

"95% confident" describes the long-run success rate of the method, not the probability this specific interval is correct

### Confusing MLE with unbiased

MLE is not automatically unbiased — the Normal variance MLE is a textbook counter-example

### Ignoring the sample-size rule for $\hat{p}$

the Normal approximation for a proportion breaks down when $n \cdot \hat{p}$ or $n \cdot \hat{q}$ is small

### Reporting a point estimate alone

a number with no interval hides how much uncertainty is actually there

# Estimation in Practice

*The same tools, applied across your specialization tracks*

## Internet of Things

### MLE

Estimating a sensor's true failure rate from field data

### CI for mean

Reporting average battery life with a confidence range

### CI for proportion

Estimating the fraction of a fleet needing firmware updates

## Cyber Security

### MLE

Estimating the average rate of intrusion attempts per hour

### CI for mean

Reporting average detection response time with uncertainty

### CI for proportion

Estimating the true rate of malicious traffic from a sample capture

## Blockchain

### MLE

Estimating the probability a transaction passes validation

### CI for mean

Reporting average confirmation time with a confidence range

### CI for proportion

Estimating the share of failed transactions network-wide

# Key Takeaways

1. A point estimate is a single best guess; an interval estimate adds a range and a confidence level, quantifying the uncertainty.
2. A good estimator is unbiased ($E(\theta) = \theta$), has minimum variance among unbiased estimators (MVUE), and is consistent (improves as n grows).
3. Maximum Likelihood Estimation picks the parameter value that makes the observed data most probable — for Binomial, Poisson, and Normal, it recovers the sample proportion, sample mean, and sample mean/variance.
4. Confidence limits for a mean or proportion all follow the same pattern: point estimate ± (critical value) × (standard error).
5. Next: Module 5 — Testing of Hypothesis, using these same estimators and standard errors to decide whether an observed difference is real or just sampling noise.