STA 130A: Mathematical Statistics: Brief Course

Lecture 25: Point Estimation and the Method of Moments

Dogyoon Song

Fall 2026, UC Davis

Agenda

Last time: Confidence intervals

- Use a sampling distribution to quantify uncertainty about a mean.

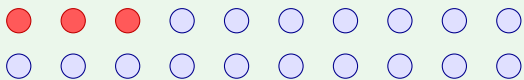
Today: Constructing and assessing point estimators

- **Goal:** Estimate an unknown parameter from a sample.
- **Construction of an estimator:** Use the method of moments.
- **Assessment of estimators:**
  - **At a fixed sample size:** Assess bias, variance, and MSE.
  - **As the sample grows:** Ask whether the estimator is consistent.

Motivation: How common are defective items?

Example

A manufacturer inspects 20 independently sampled items and finds **3 defective items**.



Red: defective

The unknown quantity is the **defect probability** p. How should we estimate it from these observations?

A natural answer is the observed fraction:

$$\hat{p}_{\text{obs}} = \frac{3}{20} = 0.15.$$

Today's questions: How can we construct estimators systematically, and how can we assess their accuracy?

Point estimation: Parameter, estimator, and estimate

Definition (Point estimator)

To estimate a fixed unknown parameter θ, choose a rule based on the sample:

$$\hat{\theta}_n = T(X_1, \ldots, X_n).$$

The **estimator** is random; its observed value is the **estimate**.

Example (The defect probability)

Let $X_i = 1$ for a defective item and $X_i = 0$ otherwise.

Parameter	p: the fixed unknown defect probability
Estimator	$\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$: the rule
Estimate	$\hat{p}_{\text{obs}} = 3/20 = 0.15$: its observed value

Our task: Construct a rule; then study its sampling distribution.

Population moments and sample moments

Definition (Moments about 0)

The kth **population moment** and **sample moment** are respectively

$$m_k = \mathbb{E}[X^k], \qquad \widehat{m}_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k.$$

Example (Observed data: $1, 2, 3, 4$)

	Population	Sample	Observed value
First moment	$\mathbb{E}[X]$	$\bar{X}_n$	2.5
Second moment	$\mathbb{E}[X^2]$	$\frac{1}{n}\sum_{i=1}^{n} X_i^2$	7.5

Recall: $\mathbb{E}[X^2] = \mathrm{Var}(X) + (\mathbb{E}[X])^2$.

Method of moments: One parameter

A recipe for constructing an estimator

1. Choose a population moment that determines the unknown parameter.
2. Match it to the corresponding sample moment.
3. Solve for the parameter and check that the solution is valid.

Often, we use the first moment:

$$m_1(\theta) = \mathbb{E}_\theta[X].$$

The moment equation for the estimator is

$$m_1(\hat{\theta}_{\mathrm{MM}}) = \bar{X}_n.$$

Our task: Solve this equation for $\hat{\theta}_{\mathrm{MM}}$.

Why match moments? By the WLLN, sample moments approach population moments as the sample size n grows.

Example: Estimating a defect probability

Example (Bernoulli model)

Let $X_i = 1$ if item i is defective and $X_i = 0$ otherwise. Model $X_1, \ldots, X_n$ as i.i.d. $\mathrm{Bernoulli}(p)$.

The model relates the unknown parameter to a population mean:

$$\mathbb{E}[X_i] = p.$$

Replace the population mean by the sample mean:

$$\hat{p}_{\mathrm{MM}} = \bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

For the observed sample, the estimate is

$$\hat{p}_{\mathrm{obs}} = \frac{3}{20} = 0.15.$$

Idea: Match a population moment to its sample counterpart.

Example: Estimating an exponential rate

Example (Waiting times)

Suppose $X_1, \ldots, X_n$ are i.i.d. waiting times with

$$X_i \sim \text{Exponential}(\lambda), \qquad f_X(x) = \lambda e^{-\lambda x}, \quad x \geq 0, \quad \lambda > 0.$$

The first population moment is

$$\mathbb{E}[X_i] = \frac{1}{\lambda}.$$

Match it to the sample mean and solve:

$$\frac{1}{\hat{\lambda}_{\text{MM}}} = \bar{X}_n \qquad \Longrightarrow \qquad \hat{\lambda}_{\text{MM}} = \frac{1}{\bar{X}_n}.$$

Interpretation: Shorter average waiting times suggest a larger rate.

Example: From waiting times to a rate estimate

Example

Four observed waiting times, in minutes, are

$$1, \quad 2, \quad 3, \quad 4.$$

Their average is

$$\bar{x} = \frac{1+2+3+4}{4} = 2.5 \text{ minutes.}$$

The method-of-moments estimate of the rate is

$$\hat{\lambda}_{\text{obs}} = \frac{1}{2.5} = 0.4 \text{ per minute.}$$

Quick check: If the average were 4 minutes, what rate would you estimate?

Answer: 1/4 per minute. Check both the formula and the units.

Why use more than one moment equation?

Example (A normal model with unknown μ and σ^2)

The first two population moments are

$$\mathbb{E}[X] = \mu, \qquad \mathbb{E}[X^2] = \mu^2 + \sigma^2.$$

The first moment determines μ, but gives no information about σ^2.

Matching both moments to their sample counterparts gives the system

$$\begin{cases} \hat{\mu}_{\mathrm{MM}} = \bar{X}_n, \\ \hat{\mu}^2_{\mathrm{MM}} + \hat{\sigma}^2_{\mathrm{MM}} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i^2. \end{cases}$$

The first equation gives $\hat{\mu}_{\mathrm{MM}}$; substituting it into the second gives $\hat{\sigma}^2_{\mathrm{MM}}$.

What if we match more moments? Extra equations may repeat information or may not all fit exactly. One approach is to choose parameters that make the overall mismatch small.

Our focus: Match one informative moment per parameter, and check for a unique, valid solution.

Example: Two moments, two normal parameters

Example

Suppose $X_1, \ldots, X_n$ are i.i.d. $N(\mu, \sigma^2)$, with $n \geq 2$, and both parameters are unknown. The first two population moments are

$$\mathbb{E}[X] = \mu, \qquad \mathbb{E}[X^2] = \mu^2 + \sigma^2.$$

The moment equations are

$$\hat{\mu}_{\mathrm{MM}} = \bar{X}_n, \qquad \hat{\mu}^2_{\mathrm{MM}} + \hat{\sigma}^2_{\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2.$$

Solving gives

$$\hat{\mu}_{\mathrm{MM}} = \bar{X}_n, \qquad \hat{\sigma}^2_{\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

Notice: Moment matching produces the denominator n.

Example: Two variance estimates from one dataset

Example (A sample of four package weights)

For the observations $498, 500, 502, 504$ grams,

$$\bar{x} = 501, \qquad \sum_{i=1}^{4}(x_i - \bar{x})^2 = 20.$$

The method-of-moments estimate of the mean is 501 grams.

For the variance, compare

$$\hat{\sigma}^2_{\mathrm{MM,obs}} = \frac{20}{4} = 5, \qquad s^2 = \frac{20}{3} \approx 6.67 \quad \text{(grams squared)}.$$

Both rules use the same data; their denominators differ.

Next question: How can we assess and compare these estimators?

We study their accuracy across repeated samples: where are the estimates centered, and how far are they from the target?

Bias and unbiasedness

Assessment at a fixed n: Is the estimator centered at its target?

Definition (Bias)

For an estimator $\hat{\theta}$ of θ, its **bias** is

$$\text{Bias}(\hat{\theta}) = \mathbb{E}[\hat{\theta}] - \theta.$$

The estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$ for every possible value of θ.

Example (Estimating a population mean)

For an i.i.d. sample with mean μ,

$$\mathbb{E}[\bar{X}_n] = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[X_i] = \mu.$$

Thus, $\bar{X}_n$ is an unbiased estimator of μ.

Interpretation: Unbiasedness concerns the center over repeated samples. An individual estimate can still be far from the parameter.

Example: Moment matching can give a biased estimator

Example (The normal variance estimator)

1. Obtain the MM estimator by matching moments. Using $\sigma^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$, replace population moments by sample moments:

$$\hat{\sigma}^2_{\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2.$$

2. Relate it to S^2: $\hat{\sigma}^2_{\mathrm{MM}} = \frac{n-1}{n}S^2$, since S^2 uses denominator $n-1$.

3. Take expectations. Using $\mathbb{E}[S^2] = \sigma^2$,

$$\mathbb{E}[\hat{\sigma}^2_{\mathrm{MM}}] = \frac{n-1}{n}\mathbb{E}[S^2] = \frac{n-1}{n}\sigma^2.$$

4. Subtract the target σ^2.

$$\mathrm{Bias}(\hat{\sigma}^2_{\mathrm{MM}}) = \mathbb{E}[\hat{\sigma}^2_{\mathrm{MM}}] - \sigma^2 = -\frac{\sigma^2}{n}.$$

For $n = 4$: Across repeated samples, the estimates average $3\sigma^2/4$, or 25% below the true variance.

Variance and mean squared error

Estimator variance measures spread around the estimator's own mean:

$$\mathrm{Var}(\hat{\theta}) = \mathbb{E}\Big[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2\Big].$$

Definition (Mean squared error)

The **mean squared error** measures average squared distance from the true parameter:

$$\mathrm{MSE}(\hat{\theta}) = \mathbb{E}\Big[(\hat{\theta} - \theta)^2\Big].$$

Two centers: Variance uses $\mathbb{E}[\hat{\theta}]$; MSE uses the target θ. The difference between these centers is the bias.

Question: How do spread and bias together determine MSE?

MSE equals variance plus squared bias

Bias–variance decomposition

$$\mathrm{MSE}(\hat{\theta}) = \mathrm{Var}(\hat{\theta}) + \left(\mathrm{Bias}(\hat{\theta})\right)^2.$$

Let $m = \mathbb{E}[\hat{\theta}]$, the estimator's mean over repeated samples.

Split the estimation error into two parts:

$$\hat{\theta} - \theta = \underbrace{(\hat{\theta} - m)}_{\text{fluctuation around its mean}} + \underbrace{(m - \theta)}_{\text{bias}}.$$

Square and average, using $(a+b)^2 = a^2 + 2ab + b^2$:

$$\mathbb{E}[(\hat{\theta} - \theta)^2] = \mathbb{E}[(\hat{\theta} - m)^2] + 2(m - \theta)\underbrace{\mathbb{E}[\hat{\theta} - m]}_{= m - m = 0} + (m - \theta)^2.$$

The cross term vanishes because fluctuations around the mean average to zero.

Consequence: For an unbiased estimator, MSE equals its variance.

Example: One observation versus the sample mean

Example (Estimating the mean package weight)

A manufacturer records n i.i.d. package weights, each with mean μ and variance σ^2. To estimate μ, compare using the first weight with averaging all the weights:

$$T_1 = X_1, \qquad T_2 = \bar{X}_n.$$

Question: Both are unbiased. Are they equally accurate?

	$T_1 = X_1$	$T_2 = \bar{X}_n$
Expectation	μ	μ
Bias	0	0
Variance	σ^2	σ^2/n
MSE	σ^2	σ^2/n

Concrete check: If $\sigma = 4$ grams and $n = 16$, the MSEs are 16 and 1 grams squared, respectively.

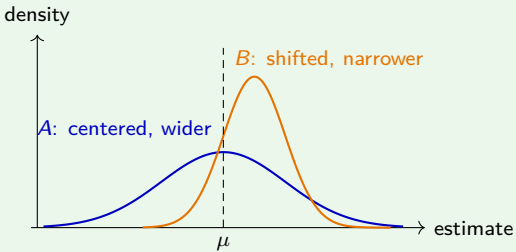
Message: Both have zero bias, so the smaller variance gives the smaller MSE.

Example: Can a biased estimator have smaller MSE?

Example (Two estimators of the same mean)

Suppose two procedures have the following sampling distributions:

$$A \sim N(\mu, 1), \qquad B \sim N(\mu + 0.5, 0.25).$$



	A	B
Bias	0	0.5
Variance	1	0.25
MSE	1	$0.25 + 0.5^2 = 0.5$

Message: Here, B's smaller variance outweighs its squared bias. An unbiased estimator need not have the smallest MSE.

Consistency: Accuracy as the sample grows

At a fixed n: Bias and MSE assess accuracy at that sample size.

As n grows: Does collecting more data make large errors increasingly unlikely?

Definition (Consistency)

A sequence of estimators $\hat{\theta}_n$ is **consistent** for θ if

$$\hat{\theta}_n \xrightarrow{P} \theta, \qquad \text{that is,} \qquad P(|\hat{\theta}_n - \theta| \geq \epsilon) \longrightarrow 0$$

for every fixed $\epsilon > 0$ and every possible value of θ.

Interpretation: Any fixed error tolerance is met with probability approaching 1 as the sample grows.

Complementary criteria: Consistency gives a long-run guarantee. It does not specify the accuracy at a particular n or how large n must be.

Example: Unbiased and consistent

Example (The method-of-moments estimator of a defect probability)

For i.i.d. $X_i \sim \text{Bernoulli}(p)$, use the sample proportion

$$\hat{p}_n = \bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i.$$

Unbiased: Its expectation equals the parameter at every n:

$$\mathbb{E}[\hat{p}_n] = \mathbb{E}[\bar{X}_n] = p.$$

Consistent: By the WLLN,

$$\hat{p}_n = \bar{X}_n \xrightarrow{P} p.$$

Its MSE also decreases as the sample grows:

$$\text{MSE}(\hat{p}_n) = \text{Var}(\hat{p}_n) = \frac{p(1-p)}{n} \longrightarrow 0.$$

Message: This MoM estimator is centered correctly and becomes more accurate.

Example: Unbiased but inconsistent

Example

Suppose $X_1, X_2, \ldots$ are i.i.d. $\mathrm{Bernoulli}(p)$. For every n, use only the first observation:

$$T_n = X_1.$$

Unbiased: Its expectation equals the parameter:

$$\mathbb{E}[T_n] = p \qquad \text{for every } n.$$

Inconsistent: At $p = 1/2$, T_n is either 0 or 1, so

$$P\left(\left|T_n - \frac{1}{2}\right| \geq \frac{1}{4}\right) = 1 \qquad \text{for every } n.$$

*** Why inconsistent?** The estimator ignores all additional observations, not benefitting from increasing n.

Message: Unbiasedness alone does not ensure consistency.

Example: Biased but consistent

Example (Revisit the normal method-of-moments variance estimator)

Biased: For an i.i.d. normal sample, we already found

$$\hat{\sigma}^2_{\mathrm{MM}} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}_n^2, \qquad \mathrm{Bias}(\hat{\sigma}^2_{\mathrm{MM}}) = -\frac{\sigma^2}{n}.$$

Consistent: Apply the WLLN to X_i and to X_i^2:

$$\bar{X}_n \xrightarrow{P} \mu, \qquad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} \mu^2 + \sigma^2.$$

Squaring and subtracting preserve these probability limits, so

$$\hat{\sigma}^2_{\mathrm{MM}} \xrightarrow{P} (\mu^2 + \sigma^2) - \mu^2 = \sigma^2.$$

Message: Consistency does not guarantee unbiasedness.

Pop-up quiz: Assessing an estimator

Question: Which statement is false?

A) An unbiased estimator has expectation equal to the parameter.

B) For an unbiased estimator, MSE equals variance.

C) Every unbiased estimator is consistent.

D) A consistent estimator can be biased at each finite sample size.

Answer: C. At $p = 1/2$, the unbiased rule $T_n = X_1$ stays $1/2$ away from its target for every n.

Two follow-up questions:

- Where is the sampling distribution centered?
- Does it concentrate near the parameter as n increases?

Wrap-up: What is our verdict on MoM?

MoM is a useful, principled way to construct estimators.

The WLLN makes sample moments approach population moments. If the chosen moments uniquely and continuously determine the parameters, this gives **consistent** MoM estimators.

MoM estimator	Unbiased?	Consistent?
Bernoulli: $\hat{p}_n = \bar{X}_n$	Yes	Yes
Normal variance: $\hat{\sigma}^2_{\mathrm{MM}}$	No	Yes

At a fixed n: Assess bias and MSE.

- MoM does not guarantee the smallest MSE.

Suggested reading: [Ric07, Chapters 8 and 4].

Next lecture: Maximum likelihood estimation.

References



John A Rice.
Mathematical statistics and data analysis, volume 9.
Thomson brooks/cole, Belmont, CA, 2007.