

Statistics for Economists II

Week 6: Method of Moments & Maximum Likelihood Estimation

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Motivation: Why Do We Need Estimation Methods?
- 2 Method of Moments (MoM)
- 3 Maximum Likelihood Estimation (MLE)
- 4 Properties of MLE
- 5 MLE vs. MoM: A Comparison
- 6 Economics Applications
- 7 Practice Problems
- 8 Key Takeaways

The Estimation Problem

Recall from previous weeks:

- We observe a random sample $X_1, \dots, X_n$ from a distribution $f(x; \theta)$

So far we have been *given* estimators (e.g., $\bar{X}$ for μ) and checked their properties. Today we learn two **systematic recipes** for *constructing* estimators from scratch.

Today's Two Methods

- ① **Method of Moments (MoM)** — simple, intuitive, always works
- ② **Maximum Likelihood Estimation (MLE)** — the gold standard of estimation

The Estimation Problem

Recall from previous weeks:

- We observe a random sample $X_1, \dots, X_n$ from a distribution $f(x; \theta)$
- The parameter θ is **unknown**—our goal is to estimate it

So far we have been *given* estimators (e.g., $\bar{X}$ for μ) and checked their properties. Today we learn two **systematic recipes** for *constructing* estimators from scratch.

Today's Two Methods

- ① **Method of Moments (MoM)** — simple, intuitive, always works
- ② **Maximum Likelihood Estimation (MLE)** — the gold standard of estimation

The Estimation Problem

Recall from previous weeks:

- We observe a random sample $X_1, \dots, X_n$ from a distribution $f(x; \theta)$
- The parameter θ is **unknown**—our goal is to estimate it
- We know *desirable properties*: unbiasedness, consistency, efficiency

So far we have been *given* estimators (e.g., $\bar{X}$ for μ) and checked their properties. Today we learn two **systematic recipes** for *constructing* estimators from scratch.

Today's Two Methods

- ➊ **Method of Moments (MoM)** — simple, intuitive, always works
- ➋ **Maximum Likelihood Estimation (MLE)** — the gold standard of estimation

The Estimation Problem

Recall from previous weeks:

- We observe a random sample $X_1, \dots, X_n$ from a distribution $f(x; \theta)$
- The parameter θ is **unknown**—our goal is to estimate it
- We know *desirable properties*: unbiasedness, consistency, efficiency
- But **how do we actually find** an estimator in the first place?

So far we have been *given* estimators (e.g., $\bar{X}$ for μ) and checked their properties. Today we learn two **systematic recipes** for *constructing* estimators from scratch.

Today's Two Methods

- ➊ **Method of Moments (MoM)** — simple, intuitive, always works
- ➋ **Maximum Likelihood Estimation (MLE)** — the gold standard of estimation

Moment Generating Functions (MGFs)

Definition

For a random variable X , the **moment generating function (MGF)** is:

$$M_X(t) = E \left[e^{tX} \right]$$

provided this expectation exists for t in some open interval around 0.

Key Property: Extract Moments by Differentiating at $t = 0$

Since $\frac{d}{dt} e^{tX} = X e^{tX}$, differentiating inside the expectation gives:

$$M'_X(0) = E[X] = \mu, \quad M''_X(0) = E[X^2] = \mu^2 + \sigma^2$$

More generally, $M_X^{(n)}(0) = E[X^n]$ — every moment is a derivative of $M_X(t)$ at zero, **no integration required**.

MoM Estimator $\neq$ MGF

MGF: $M_X(t) = E[e^{tX}]$

- A function of t belonging to a **distribution**
- **Theoretical** — depends on θ , not on data
- Gives population moments:
 $M'_X(0) = E[X]$

MoM Estimator: $\hat{\theta}_{\text{MoM}}$

- A function of the **sample** $X_1, \dots, X_n$
- **Empirical** — changes with every new sample
- Obtained by solving $\mu'_k(\theta) = m'_k$

Key Distinction

The MGF is a **tool for Step 1** of MoM: it gives $E[X^k]$ as a function of θ without direct integration. It is **not** the estimator. The estimator is the final $\hat{\theta} = h(\bar{X})$ obtained after solving the moment equations.

The Big Idea Behind MoM

Population moments are **functions of parameters**. Sample moments are **computable from data**. **Set them equal** and **solve for the parameters**.

Population Moments: functions of θ

$$\mu'_k = \mathbb{E}(X^k) = \int x^k f(x; \theta) dx \quad \Rightarrow \quad \mu'_1(\theta), \mu'_2(\theta), \dots$$

Sample Moments: computable from data

$$m'_k = \frac{1}{n} \sum_{i=1}^n X_i^k \quad \Rightarrow \quad m'_1 = \bar{X}, \quad m'_2 = \frac{1}{n} \sum X_i^2, \dots$$

MoM: Step-by-Step Procedure

Suppose $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ has p unknown parameters.

Step 1: Compute the first p population moments as functions of θ :

$$\mu'_k = \mathbb{E}(X^k) = g_k(\theta_1, \dots, \theta_p), \quad k = 1, 2, \dots, p$$

Step 2: Set population moments equal to sample moments:

$$g_k(\theta_1, \dots, \theta_p) = m'_k = \frac{1}{n} \sum_{i=1}^n X_i^k, \quad k = 1, 2, \dots, p$$

Step 3: Solve the system of p equations for p unknowns $\hat{\theta}_1, \dots, \hat{\theta}_p$.

One parameter $\Rightarrow$ one equation (first moment). Two parameters $\Rightarrow$ two equations (first and second moments). It is like matching the “summary statistics” of your model to what you see in the data.

MoM Example 1: Exponential Distribution

Let $X_1, \dots, X_n$ be a random sample from $f(x; \theta) = \theta e^{-\theta x}$, $x > 0$. Find $\hat{\theta}_{\text{MoM}}$.

Step 1: Find $E(X)$ via the MGF.

$$\begin{aligned} M_X(t) &= E(e^{tX}) = \int_0^{\infty} e^{tx} \cdot \theta e^{-\theta x} dx = \theta \int_0^{\infty} e^{-(\theta-t)x} dx \\ &= \theta \left[\frac{-1}{\theta-t} e^{-(\theta-t)x} \right]_0^{\infty} = \frac{\theta}{\theta-t}, \quad t < \theta \end{aligned}$$

Differentiate and evaluate at $t = 0$:

$$M'_X(t) = \frac{\theta}{(\theta-t)^2} \implies E(X) = M'_X(0) = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

Step 2: Sample moment: $m_1 = \bar{X}$ **Step 3:** Equate and solve:

$$\frac{1}{\theta} = \bar{X} \implies \boxed{\hat{\theta}_1 = \frac{1}{\bar{X}}}$$

MoM Example 2: Normal Distribution — with Real Data

Problem: Monthly book expenditures (TL) of 10 students :

5, 15, 20, 40, 40, 45, 50, 50, 50, 55

Assume $X \sim N(\mu, \sigma^2)$. Estimate μ and σ^2 by MoM.

Step 1: Population moments: $\mathbb{E}(X) = \mu$, $\mathbb{E}(X^2) = \sigma^2 + \mu^2$.

Step 2: Sample moments:

$$m_1 = \bar{x} = \frac{370}{10} = 37, \quad m_2 = \frac{1}{n} \sum_{i=1}^n x_i^2 = \frac{16,400}{10} = 1640$$

Step 3: Solve $\mu = m_1$ and $\sigma^2 + \mu^2 = m_2$:

$$\hat{\mu}_{\text{MoM}} = 37,$$

$$\hat{\sigma}_{\text{MoM}}^2 = 1640 - 37^2 = 271$$

Note

$\hat{\sigma}_{\text{MoM}}^2$ divides by n (not $n - 1$) — **biased** downward, but consistent.

MoM Example 3: A Non-Standard Distribution

Problem (Question 3.8): Let $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} f(y | \theta)$ where

$$f(y | \theta) = \frac{2}{\theta^2}(\theta - y), \quad 0 \leq y \leq \theta.$$

Find $\hat{\theta}_{\text{MoM}}$.

Step 1: Compute $\mathbb{E}[Y]$ by integration:

$$\mathbb{E}[Y] = \int_0^\theta y \cdot \frac{2}{\theta^2}(\theta - y) dy = \frac{2}{\theta^2} \left[\frac{\theta y^2}{2} - \frac{y^3}{3} \right]_0^\theta = \frac{2}{\theta^2} \cdot \frac{\theta^3}{6} = \frac{\theta}{3}$$

Steps 2 & 3: Set $\mathbb{E}[Y] = \bar{Y}$ and solve:

$$\frac{\theta}{3} = \bar{Y} \implies \boxed{\hat{\theta}_{\text{MoM}} = 3\bar{Y}}$$

MoM works for *any* distribution: compute $\mathbb{E}[Y]$ as a function of θ , set equal to $\bar{Y}$, solve. No optimization needed.

MoM: Advantages and Disadvantages

Advantages

- **Simple:** solve equations—no optimization needed
- **Always works:** moments exist under mild conditions
- **Consistent:** by LLN, $m'_k \xrightarrow{P} \mu'_k$
- Good starting point for advanced methods

Disadvantages

- **Not always efficient:** may not achieve minimum variance
- May yield **inadmissible estimates** (e.g., $\hat{\sigma}^2 < 0$)
- **Not unique:** different moment conditions, different estimators
- Ignores the *shape* of the distribution

The Red Balls Story: Maximum Likelihood Estimation (MLE) Intuition

Imagine a box with 10 balls — some red, some blue. You do not know how many are red.

You draw 3 balls **with replacement** and observe:
Red, Red, Blue.

Which scenario is most likely?

- If 1 red ball: $\mathbb{P} = (0.1)^2(0.9) = 0.009$
- If 3 red balls: $\mathbb{P} = (0.3)^2(0.7) = 0.063$
- If 5 red balls: $\mathbb{P} = (0.5)^2(0.5) = 0.125$
- If 7 red balls: $\mathbb{P} = (0.7)^2(0.3) = 0.147$ ←
max
- If 9 red balls: $\mathbb{P} = (0.9)^2(0.1) = 0.081$

MLE is **detective work**: given the evidence (data), what parameter value best explains what we observed?

The maximum likelihood estimate is the value of θ that makes the observed data **most probable**.

We pick $\hat{p} = 0.7$ because 7 red balls makes seeing “Red, Red, Blue” most likely.

The Red Balls Story: Exact MLE

In the red balls example, the exact MLE is:

Let p = probability of red. We observed: Red, Red, Blue.

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

Take the derivative and set it to zero:

$$\frac{dL}{dp} = 2p(1 - p) - p^2 = p(2 - 3p) = 0$$

Solutions: $p = 0$ (minimum) or $\hat{p} = 2/3 \approx 0.667$.

The MLE is $\hat{p} = 2/3$, which equals the **sample proportion** of red balls (2 reds out of 3 draws). This is not a coincidence. The MLE for a Bernoulli proportion is always the sample proportion.

The Likelihood Function

Definition

Given i.i.d. observations $x_1, \dots, x_n$ from $f(x; \theta)$, the **likelihood function** is:

$$L(\theta) = L(\theta \mid x_1, \dots, x_n) = \prod_{i=1}^n f(x_i; \theta)$$

Key Conceptual Shift

The pdf $f(x; \theta)$ treats θ as **fixed** and x as **variable**.

The likelihood $L(\theta)$ treats the data as **fixed** and θ as **variable**.

Same formula, different perspective

Think of the likelihood as asking: “For each possible value of θ , how probable is it that we would have observed *exactly this data*?” We then choose the θ that answers “most probable.”

Why Take Logs? The Log-Likelihood

The Log-Likelihood Function

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

Why is this better?

- ➊ **Products become sums:** much easier to differentiate
- ➋ **Numerical stability:** products of many small numbers $\rightarrow$ underflow; sums of logs avoid this
- ➌ **Same maximizer:** $\ln$ is strictly increasing, so $\hat{\theta}$ that maximizes $\ell(\theta)$ also maximizes $L(\theta)$

Imagine multiplying 100 probabilities, each around 0.01. The product is roughly 10^{-200} , which is a hard to store number but $\sum \ln(0.01) = -200 \ln(10) \approx -460.5$, which is perfectly manageable.

MLE: Step-by-Step Procedure

Step 1: Write down the likelihood function:

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

Step 2: Take the log-likelihood:

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

Step 3: First-order condition: differentiate and set to zero:

$$\frac{\partial \ell}{\partial \theta} = 0 \quad (\text{the score equation})$$

Step 4: Solve for $\hat{\theta}_{\text{MLE}}$.

Step 5: Verify it is a maximum (check second-order condition: $\frac{\partial^2 \ell}{\partial \theta^2} \Big|_{\hat{\theta}} < 0$).

MLE Example 1: Exponential Distribution

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

Step 1–2: Log-likelihood:

$$\ell(\lambda) = \sum_{i=1}^n \ln(\lambda e^{-\lambda x_i}) = n \ln \lambda - \lambda \sum_{i=1}^n x_i$$

Step 3: Score equation:

$$\frac{d\ell}{d\lambda} = \frac{n}{\lambda} - \sum_{i=1}^n x_i = 0$$

Step 4: Solve:

$$\hat{\lambda}_{\text{MLE}} = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{X}}$$

Step 5: $\frac{d^2\ell}{d\lambda^2} = -\frac{n}{\lambda^2} < 0$ ✓ (confirmed maximum)

MLE Example 2: Poisson Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$ with $p(x; \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0, 1, 2, \dots$

Log-likelihood:

$$\ell(\lambda) = \sum_{i=1}^n [-\lambda + x_i \ln \lambda - \ln(x_i!)] = -n\lambda + \ln \lambda \sum_{i=1}^n x_i - \sum_{i=1}^n \ln(x_i!)$$

Score equation:

$$\frac{d\ell}{d\lambda} = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i = 0$$

Solution:

$$\hat{\lambda}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X}$$

If X counts the number of customers arriving at a bank branch per hour, then $\hat{\lambda} = \bar{X}$ (the sample average arrival rate) is the MLE.

MLE Example 3: Bernoulli / Binomial

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$. Each $X_i \in \{0, 1\}$.

Log-likelihood:

$$\ell(p) = \sum_{i=1}^n [x_i \ln p + (1 - x_i) \ln(1 - p)] = \left(\sum x_i\right) \ln p + \left(n - \sum x_i\right) \ln(1 - p)$$

Score equation:

$$\frac{d\ell}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1 - p} = 0$$

Solve: Cross-multiply and simplify:

$$\hat{p}_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} = \hat{p} \quad (\text{sample proportion})$$

Estimating the probability of loan default. If 23 out of 200 loans default, then $\hat{p}_{\text{MLE}} = 23/200 = 0.115$.

MLE Example 4: Normal Distribution

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$. Two unknowns: μ and σ^2 .

Log-likelihood:

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

FOC w.r.t. μ :

$$\frac{\partial \ell}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu) = 0 \implies \boxed{\hat{\mu}_{\text{MLE}} = \bar{X}}$$

FOC w.r.t. σ^2 :

$$\frac{\partial \ell}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2} \sum_{i=1}^n (x_i - \mu)^2 = 0 \implies \boxed{\hat{\sigma}_{\text{MLE}}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}$$

Important

Just like MoM, the MLE for σ^2 divides by n (not $n-1$), so it is biased. But it is consistent and asymptotically efficient.

Economic Application: Estimating Wage Distribution Parameters

Setting: Labor economists often model log-wages as normally distributed:

$$\ln W_i \sim N(\mu, \sigma^2)$$

Data: A random sample of $n = 500$ workers gives:

$$\overline{\ln W} = 8.42, \quad \frac{1}{n} \sum_{i=1}^n (\ln W_i - \overline{\ln W})^2 = 0.36$$

MLE estimates:

$$\hat{\mu}_{\text{MLE}} = 8.42, \quad \hat{\sigma}_{\text{MLE}}^2 = 0.36$$

These estimates fully characterize the wage distribution. From them, we can estimate:

- Median wage: $e^{\hat{\mu}} = e^{8.42} \approx 4,527$ TL (For a lognormal distribution, the median of W is e^{μ})
- Mean wage: $e^{\hat{\mu} + \hat{\sigma}^2/2} = e^{8.60} \approx 5,420$ TL (For a lognormal distribution, the mean of W is $e^{\mu + \sigma^2/2}$)
- Wage inequality (Gini coefficient depends on σ)

The Invariance Property of MLE

Theorem (Invariance Property)

If $\hat{\theta}_{MLE}$ is the MLE of θ , then for any function $g(\theta)$:

$$\widehat{g(\theta)}_{MLE} = g(\hat{\theta}_{MLE})$$

This is remarkably useful:

- MLE of λ in Exponential is $1/\bar{X} \Rightarrow$ MLE of mean $1/\lambda$ is $\bar{X}$
- MLE of σ^2 in Normal is $\hat{\sigma}^2 \Rightarrow$ MLE of σ is $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$
- MLE of p in Bernoulli is $\hat{p} \Rightarrow$ MLE of odds $p/(1-p)$ is $\hat{p}/(1-\hat{p})$

Economics example: If you estimate demand elasticity ε via MLE, you automatically get the MLE of revenue impact $1 + 1/\varepsilon$ for free—just plug in. No need to re-derive anything.

Large-Sample Properties of MLE

Under regularity conditions, MLEs enjoy remarkable asymptotic properties:

1. Consistency

$$\hat{\theta}_{\text{MLE}} \xrightarrow{P} \theta_0 \quad (\text{converges to the true value as } n \rightarrow \infty)$$

2. Asymptotic Normality

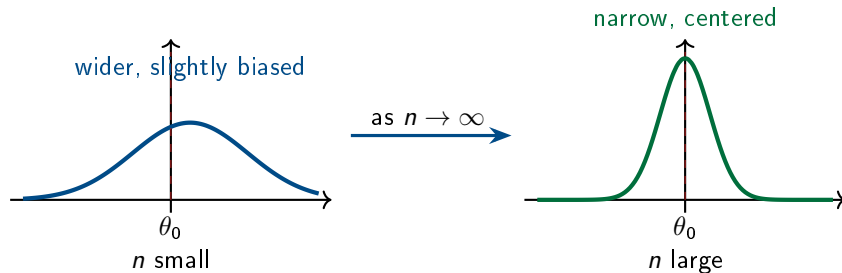
$$\sqrt{n}(\hat{\theta}_{\text{MLE}} - \theta_0) \xrightarrow{D} N\left(0, \frac{1}{I(\theta_0)}\right)$$

where $I(\theta)$ is the **Fisher information**: $I(\theta) = -\mathbb{E}\left[\frac{\partial^2 \ln f}{\partial \theta^2}\right]$

3. Asymptotic Efficiency

The MLE achieves the **Cramér–Rao lower bound** asymptotically—no consistent estimator can do better in large samples.

What Do These Properties Mean?



As the sample grows: the MLE **centers** on the truth (consistency), becomes **approximately normal** (asymptotic normality), and has the **smallest possible variance** (efficiency). This is why MLE is the workhorse of modern econometrics.

Application: Probability of Loan Default

Setting: A bank observes $n = 1,000$ loans. Each loan either defaults ($X_i = 1$) or not ($X_i = 0$).
Observed: 85 defaults.

Model: $X_i \stackrel{\text{iid}}{\sim} \text{Bernoulli}(p)$

MLE:

$$\hat{p}_{\text{MLE}} = \frac{85}{1000} = 0.085$$

Asymptotic distribution (from MLE theory):

$$\hat{p} \stackrel{\text{approx}}{\sim} N\left(p, \frac{p(1-p)}{n}\right) \implies \text{SE}(\hat{p}) = \sqrt{\frac{0.085 \times 0.915}{1000}} \approx 0.0088$$

This standard error tells the bank how precisely it knows the default rate. With $\text{SE} \approx 0.9\%$, the bank can be quite confident in its estimate—critical for setting loan loss provisions and capital requirements.

Application: Customer Arrival Rates

Setting: An economist studies foot traffic at a retail store. Over $n = 30$ one-hour periods, the number of customers arriving is recorded. The data gives $\bar{x} = 12.3$ customers/hour.

Model: $X_i \stackrel{\text{iid}}{\sim} \text{Poisson}(\lambda)$

MLE: $\hat{\lambda} = \bar{X} = 12.3$ customers/hour

By invariance:

- MLE of $\mathbb{P}(X = 0) = e^{-\hat{\lambda}} = e^{-12.3} \approx 0.0000045$ (probability of an empty hour)
- MLE of $\mathbb{P}(X > 20) = \sum_{k=21}^{\infty} \frac{e^{-12.3}(12.3)^k}{k!} \approx 0.0128$ (probability of a very busy hour)

Staffing decisions, inventory management, and store layout can all be optimized once we have $\hat{\lambda}$. The invariance property lets us estimate *any* function of λ without re-doing the MLE.

Practice Problem 1: MoM for Exponential

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$ with $f(x; \lambda) = \lambda e^{-\lambda x}$, $x > 0$. Find the MoM estimator of λ .

Practice Problem 1: MoM for Exponential

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Exp}(\lambda)$ with $f(x; \lambda) = \lambda e^{-\lambda x}$, $x > 0$. Find the MoM estimator of λ .

Solution:

Step 1: First population moment: $\mu'_1 = \mathbb{E}(X) = 1/\lambda$

Step 2: Set equal to first sample moment:

$$\frac{1}{\lambda} = \bar{X}$$

Step 3: Solve:

$$\hat{\lambda}_{\text{MoM}} = \frac{1}{\bar{X}}$$

This happens to coincide with the MLE. For the exponential distribution, MoM and MLE give the **same estimator**. This occurs whenever the sufficient statistic is a simple function of the sample mean.

Practice Problem 2: MLE for Geometric Distribution

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Geometric}(p)$ with $\mathbb{P}(X = k) = (1 - p)^{k-1}p$, $k = 1, 2, \dots$. Find the MLE of p .

Practice Problem 2: MLE for Geometric Distribution

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Geometric}(p)$ with $\mathbb{P}(X = k) = (1 - p)^{k-1}p$, $k = 1, 2, \dots$. Find the MLE of p .

Solution:

Log-likelihood:

$$\ell(p) = \sum_{i=1}^n [(x_i - 1) \ln(1 - p) + \ln p] = \left(\sum x_i - n \right) \ln(1 - p) + n \ln p$$

Score equation:

$$\frac{d\ell}{dp} = -\frac{\sum x_i - n}{1 - p} + \frac{n}{p} = 0$$

Solve: $np(1 - p)^{-1} = n \cdot p^{-1} \cdot (\sum x_i - n)$... simplifying:

$$\hat{p}_{\text{MLE}} = \frac{1}{\bar{X}} = \frac{n}{\sum_{i=1}^n x_i}$$

Since $\mathbb{E}(X) = 1/p$, this makes intuitive sense: $\hat{p} = 1/\bar{X}$.

Practice Problem 3: MLE for $\text{Uniform}(0, \theta)$

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find both the MoM and MLE estimators.

Practice Problem 3: MLE for Uniform(0, θ)

Problem: Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} \text{Uniform}(0, \theta)$. Find both the MoM and MLE estimators.

MoM: $\mathbb{E}(X) = \theta/2 = \bar{X}$, so $\hat{\theta}_{\text{MoM}} = 2\bar{X}$

MLE: The likelihood is:

$$L(\theta) = \prod_{i=1}^n \frac{1}{\theta} 1_{[0, \theta]}(x_i) = \frac{1}{\theta^n} 1\{\theta \geq x_{(n)}\}$$

where $x_{(n)} = \max(x_1, \dots, x_n)$.

For $\theta \geq x_{(n)}$, $L(\theta) = 1/\theta^n$ is **decreasing** in θ . So maximize by choosing θ as small as possible:

$$\hat{\theta}_{\text{MLE}} = X_{(n)} = \max(X_1, \dots, X_n)$$

Note

Here MoM $\neq$ MLE. The MLE uses the sample maximum, which is more efficient. This is a case where calculus-based FOC does not apply—the maximum is at a **boundary**.

Practice Problem 4: MLE with Economic Data

Problem: A development economist models the duration (in years) until a micro-enterprise fails as $X \sim \text{Exp}(\lambda)$. A sample of $n = 50$ enterprises gives $\sum x_i = 175$ years.

- (a) Find the MLE of λ (the failure rate).
- (b) Find the MLE of the median survival time.
- (c) Estimate the probability that an enterprise survives more than 5 years.

Practice Problem 4: MLE with Economic Data

Problem: A development economist models the duration (in years) until a micro-enterprise fails as $X \sim \text{Exp}(\lambda)$. A sample of $n = 50$ enterprises gives $\sum x_i = 175$ years.

- (a) Find the MLE of λ (the failure rate).
- (b) Find the MLE of the median survival time.
- (c) Estimate the probability that an enterprise survives more than 5 years.

Solutions:

- (a) $\hat{\lambda} = n / \sum x_i = 50 / 175 = 2/7 \approx 0.286$ failures per year
- (b) Median of $\text{Exp}(\lambda)$ is $\ln 2 / \lambda$. By invariance: $\widehat{\text{median}} = \ln 2 / \hat{\lambda} = \frac{175 \ln 2}{50} = 3.5 \ln 2 \approx 2.43$ years
- (c) $\mathbb{P}(X > 5) = e^{-\lambda \cdot 5}$. By invariance: $\widehat{\mathbb{P}(X > 5)} = e^{-\hat{\lambda} \cdot 5} = e^{-10/7} \approx 0.239$

About 24% of enterprises survive past 5 years. This estimate directly informs microcredit lending policies and loan term structures.

Key Takeaways

- ➊ **Method of Moments:** Equate population moments to sample moments and solve. Simple and always works, but may not be efficient.
- ➋ **Maximum Likelihood:** Choose parameter values that make the observed data most probable. Requires calculus (maximizing $\ell(\theta)$), but is the gold standard.
- ➌ **Log-likelihood:** Always work with $\ell(\theta) = \ln L(\theta)$. Products become sums, making differentiation straightforward.
- ➍ **MLE properties:** Consistent, asymptotically normal, asymptotically efficient, and invariant under reparametrization.
- ➎ **In many common cases, MoM = MLE:** They differ when the sufficient statistic is not a simple function of sample moments.
- ➏ **Economic relevance:** MLE underpins regression analysis, discrete choice models, duration models, and virtually all of modern econometrics.

Further Reading

- Wackerly, Mendenhall & Scheaffer, Ch. 9: Properties of Point Estimators and Methods of Estimation
- Hogg, McKean & Craig, Ch. 6: Maximum Likelihood Methods
- Casella & Berger, Ch. 7: Point Estimation (for a more advanced treatment)
- Wooldridge, *Introductory Econometrics*, Appendix C: Maximum Likelihood Estimation

Coming Up Next Week

Week 7: Interval Estimation and Confidence Intervals. We will use the asymptotic normality of MLE to build confidence intervals—quantifying the *precision* of our estimates.