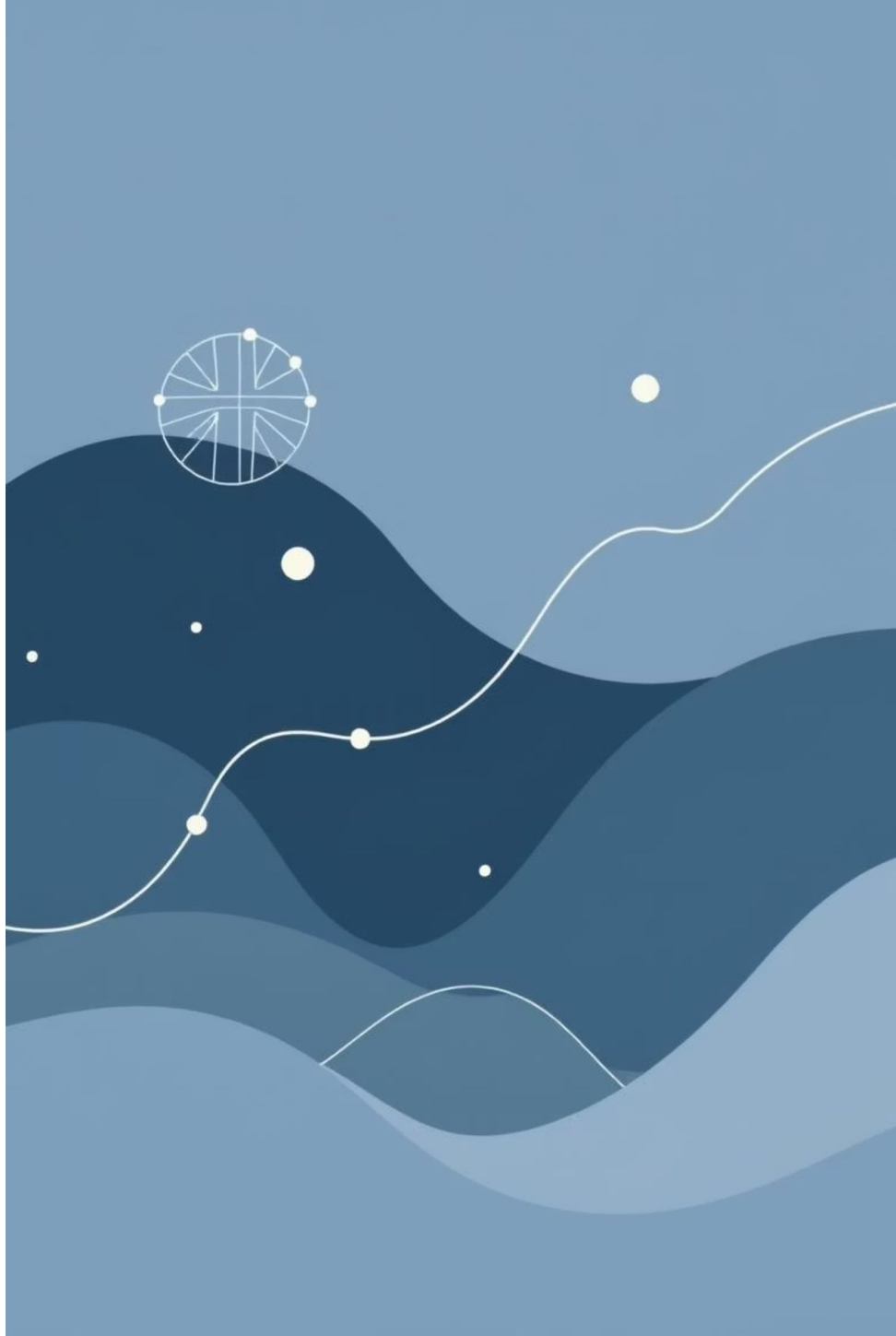
Maximum Likelihood Estimation & Parameter Estimation

Advanced Techniques for Pattern Recognition

By: Group 5 and 6

Subject: Advanced Pattern Recognition



Probability vs. Likelihood: A Fundamental Distinction

In the world of statistics and data science, understanding the difference between probability and likelihood is crucial. While often used interchangeably in everyday language, these two concepts hold distinct meanings and applications. This presentation will clarify these differences, providing a solid foundation for statistical inference.

Chapter 1

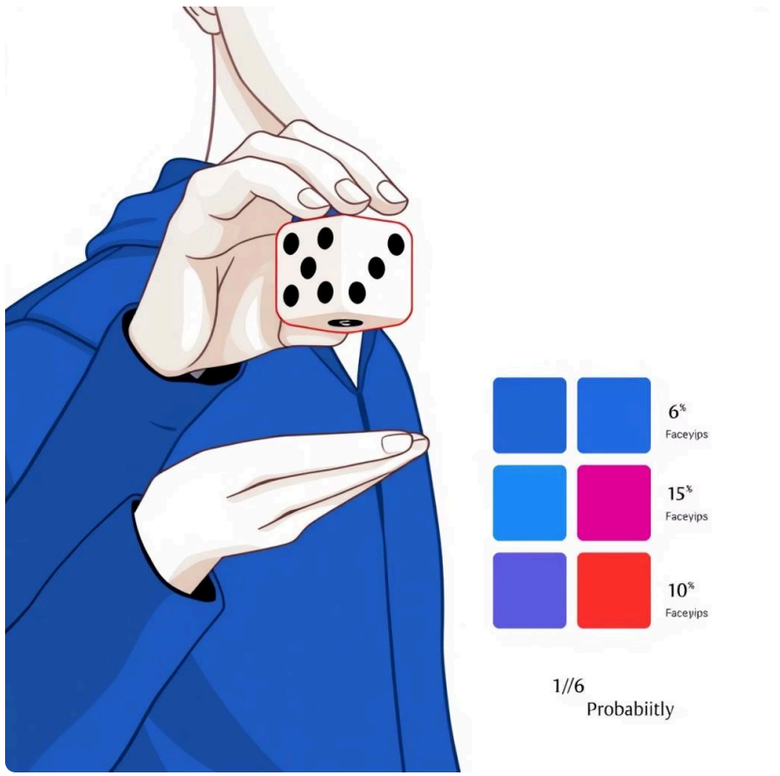
What is Probability?

Probability is a forward-looking concept. It quantifies the chance of a future event occurring, assuming you have complete knowledge of the underlying system or model.

Think of it as **predicting data** based on a known process. When we talk about probability, we're asking what outcomes are expected given a specific set of circumstances.

Key Idea: Fixed Model, Varying Outcomes

- **Given:** A fixed, known model (e.g., a fair coin with P(Heads) = 0.5).
- **Question:** What is the chance of observing a specific outcome (e.g., getting 3 heads in a row)?
- **Formula View:** P(data | model) - "The probability of the data, given the model."



Example: Rolling a Die

If you have a fair six-sided die (the **model is known**), the **probability** of rolling a 4 is exactly 1/6. You are predicting the outcome before you roll. This is a classic example of probability in action: you understand the system, and you're predicting its behavior.

An important characteristic of probabilities is that the sum of probabilities for all possible outcomes of an event always equals 1.

Chapter 2

What is Likelihood?

Example: The Mystery Coin

Imagine you flip a coin 10 times and observe 7 heads and 3 tails (the **data is known**). The **likelihood** of the coin being fair (P(Heads)=0.5) is different from the likelihood of it being biased (e.g., P(Heads)=0.7). We use the observed data to evaluate which model is a more plausible explanation.

Likelihood is a backward-looking concept. It comes into play *after* you've collected data, and it helps you assess how plausible a particular model or set of parameters is for explaining that observed data.

Think of it as **evaluating a hypothesis** based on empirical evidence. We're asking: "Given what I've seen, which explanation for this data is most credible?"

Key Idea: Fixed Data, Varying Models

- **Given:** Observed data (e.g., you flipped a coin 10 times and got 7 heads).
- **Question:** How well does a specific model (e.g., "the coin is fair") explain this data?
- **Formula View:** L(model | data) - "The likelihood of the model, given the data."

It's crucial to remember that likelihood values are used for comparison; they are **not** probabilities themselves and therefore do not sum to 1.

Parameter Estimation

Parametric Models: What Are the Parameters?

def Most random variables we've seen thus far are **parametric models**:

$$\text{Distribution} = \text{model} + \text{parameter } \theta$$

ex The distribution $\text{Ber}(0.2) \longrightarrow$ model is Bernoulli, parameter is $\theta = 0.2$.

For each of the distributions below, what is the parameter θ?

1. $\text{Ber}(p)$ $\quad \theta = p$
2. $\text{Poi}(\lambda)$ $\quad \theta = \lambda$
3. $\text{Uni}(\alpha, \beta)$ $\quad \theta = (\alpha, \beta)$
4. $\mathcal{N}(\mu, \sigma^2)$ $\quad \theta = (\mu, \sigma^2)$
5. $Y = mX + b$ $\quad \theta = (m, b)$

Parametric Models: What Are the Parameters?

In the real world, we don't know the true parameters.

- But we **observe data**: # times coin comes up heads, # per-minutes requests while using RydeShare, # visitors to website per day, strength of cell phone signal

def **estimator** $\hat{\theta}$: a **random variable** estimating the true parameter θ.

In parameter estimation,

We'll often rely on just **point estimates** — i.e., the best single value

- Provides an understanding of why data might look the way it does
- Can make future **predictions** using that model
- Can run simulations to generate more data

Maximum Likelihood Estimator

Defining Likelihood: Bernoulli

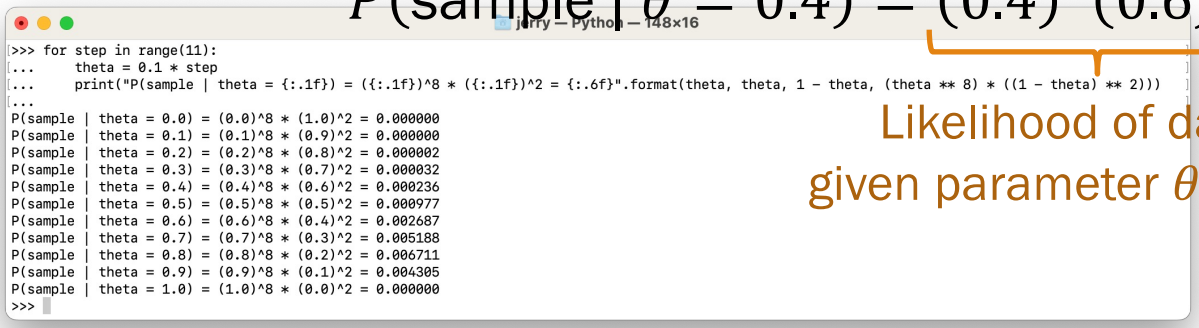
Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- X_i was drawn from distribution $F \sim \text{Ber}(\theta)$ with unknown parameter θ.
- Observed sample:

$$[1, 0, 1, 1, 1, 1, 0, 1, 1, 1] \qquad (n = 10)$$

How likely is this sample if, say, $\theta = 0.4$?

$$P(\text{sample} \mid \theta = 0.4) = \underbrace{(0.4)^8(0.6)^2}_{\text{Likelihood of data given parameter } \theta = 0.4} = 0.000236$$



Is there a better choice for θ?

Defining Likelihood

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- X_i was drawn from a distribution with density function $f(X_i|\theta)$.
- Sample: $(X_1, X_2, \ldots, X_n)$

Note: We always use $f(X_i|\theta)$, even when X_i is discrete, just so we have a uniform notation for both discrete and continuous.

Likelihood question:

How likely is the sample $(X_1, X_2, \ldots, X_n)$ given the parameter θ?

Likelihood function, $L(\theta)$:

$$L(\theta) = f(X_1, X_2, \ldots, X_n|\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

This is just a product, since the X_i are iid.

Defining Likelihood and Maximizing It

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$, drawn from a distribution $f(X_i|\theta)$.

def The **Maximum Likelihood Estimator (MLE)** of θ is the value of θ that maximizes $L(\theta)$.

$$\theta_{MLE} = \arg\max_{\theta} L(\theta)$$

the argument θ that maximizes $L(\theta)$

Likelihood of your sample

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

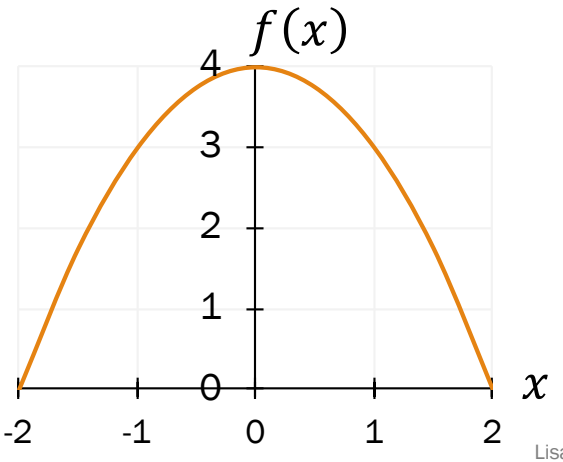
Remember: for continuous X_i, $f(X_i|\theta)$ is PDF, and for discrete X_i, $f(X_i|\theta)$ is PMF

New Function: **arg max**

$$\arg\max_x f(x)$$

The argument x that maximizes the function $f(x)$.

Let $f(x) = -x^2 + 4$, where $-2 < x < 2$.



1. $\max_x f(x)$?

$= 4$

2. $\arg\max_x f(x)$?

$= 0$

Remember: The value of x that maximizes $f(x)$ also maximizes $\log f(x)$. That's because the logarithm is a strictly increasing function for increasing values of x.

Why is this important? Because sums are easier to differentiate than products.

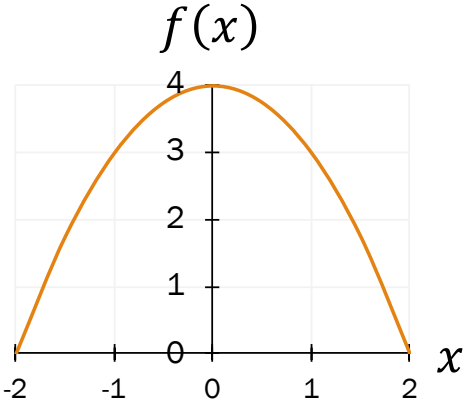
New Function: **arg max**

$$\hat{x} = \arg\max_{x} f(x)$$

Let $f(x) = -x^2 + 4$, where $-2 < x < 2$.

Differentiate wrt arg max's argument

$$\frac{d}{dx} f(x) = \frac{d}{dx}(x^2 + 4) = 2x$$



Set to 0 and solve

$$2x = 0 \quad \Rightarrow \quad \hat{x} = 0$$

Make sure $\hat{x}$ is a maximum

- Check $f(\hat{x} \pm \epsilon) < f(\hat{x})$
- We'll ignore this and not require you do this

Defining Log Likelihood and Maximizing It

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$, drawn from a distribution $f(X_i|\theta)$.

$$L(\theta) = \prod_{i=1}^{n} f(X_i|\theta)$$

θ_{MLE} maximizes the likelihood of our sample, $L(\theta)$:

$$\theta_{MLE} = \arg\max_{\theta} L(\theta)$$

θ_{MLE} also maximizes the **log-likelihood function, $LL(\theta)$**:

$$\theta_{MLE} = \arg\max_{\theta} LL(\theta)$$

$$LL(\theta) = \log L(\theta) = \log\left(\prod_{i=1}^{n} f(X_i|\theta)\right) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

$LL(\theta)$ is often easier to differentiate than $L(\theta)$.

MLE: Poisson

Maximum Likelihood Estimate: Recipe

General approach for finding θ_{MLE}, the MLE of θ:

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|\theta)$$

2. Differentiate $LL(\theta)$ wrt θ

$$\frac{\partial LL(\theta)}{\partial \theta}$$

3. Solve resulting equation

To maximize: $\frac{\partial LL(\theta)}{\partial \theta} = 0$

algebra or computer

$LL(\theta)$ is often easier to differentiate than $L(\theta)$.

Maximum Likelihood Estimate: Poisson

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = \lambda_{MLE}$?

- Let $X_i \sim \text{Poi}(\lambda)$
- PMF: $f(X_i|\lambda) = \dfrac{e^{-\lambda}\lambda^{X_i}}{X_i!}$

1. Determine formula for $LL(\theta)$

differentiating PMF directly is tough, but differentiating logarithm of it? less so

$$LL(\theta) = \sum_{i=1}^{n} \log\left(\frac{e^{-\lambda}\lambda^{X_i}}{X_i!}\right) = \sum_{i=1}^{n} (-\lambda \log e + X_i \log \lambda - \log X_i!)$$

$$= -n\lambda + \log(\lambda)\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \log(X_i!) \quad \text{(assume natural log)}$$

2. Differentiate $LL(\theta)$ wrt θ, set to 0

$$\frac{\partial LL(\theta)}{\partial \lambda} = ?$$

A. $-n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i + n\log\lambda - \sum_{i=1}^{n} \frac{1}{X_i!} \cdot \frac{\partial X_i!}{\partial X_i}$

B. $-n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i$

C. Stop trying

Maximum Likelihood Estimate: Poisson

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = \lambda_{MLE}$?

- Let $X_i \sim \text{Poi}(\lambda)$.
- PMF: $f(X_i|\lambda) = \dfrac{e^{-\lambda}\lambda^{X_i}}{X_i!}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log\left(\frac{e^{-\lambda}\lambda^{X_i}}{X_i!}\right) = \sum_{i=1}^{n} (-\lambda \log e + X_i \log \lambda - \log X_i!)$$

$$= -n\lambda + \log(\lambda)\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \log(X_i!)$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$\frac{\partial LL(\theta)}{\partial \lambda} = -n + \frac{1}{\lambda}\sum_{i=1}^{n} X_i = 0$$

3. Solve resulting equation

$$\lambda_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

MLE of the Poisson parameter, λ_{MLE}, is the **sample mean**, $\bar{X}$.
Is it always the sample mean? No!

MLE: Gaussian

Maximum Likelihood Estimate: Gaussian

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

$$f(X_i|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ wrt each θ, set to 0
3. Solve resulting equations

$$LL(\theta) = \sum_{i=1}^{n} \log\left(\frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}\right) = \sum_{i=1}^{n} \left[-\log\left(\sqrt{2\pi}\sigma\right) - (X_i-\mu)^2/(2\sigma^2)\right]$$

$$= -\sum_{i=1}^{n} \log\left(\sqrt{2\pi}\sigma\right) - \sum_{i=1}^{n} [(X_i-\mu)^2/(2\sigma^2)]$$

Maximum Likelihood Estimate: Gaussian

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

$$f(X_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ wrt each θ, set to 0
3. Solve resulting equations

$$LL(\theta) = -\sum_{i=1}^{n} \log\left(\sqrt{2\pi}\sigma\right) - \sum_{i=1}^{n} [(X_i-\mu)^2/(2\sigma^2)]$$

with respect to μ

$$\frac{\partial LL(\theta)}{\partial \mu} = \sum_{i=1}^{n} [2(X_i-\mu)/(2\sigma^2)]$$

$$= \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0$$

with respect to σ

$$\frac{\partial LL(\theta)}{\partial \sigma} = -\sum_{i=1}^{n} \frac{1}{\sigma} + \sum_{i=1}^{n} 2(X_i-\mu)^2/(2\sigma^3)$$

$$= -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 = 0$$

Maximum Likelihood Estimate: Gaussian

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

- Let $X_i \sim \mathcal{N}(\mu, \sigma^2)$.

$$f(X_i|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(X_i-\mu)^2/(2\sigma^2)}$$

What is $\theta_{MLE} = (\mu_{MLE}, \sigma^2_{MLE})$?

3. Solve resulting equations

Two equations, two unknowns:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i-\mu) = 0 \qquad -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = 0$$

First, solve for μ_{MLE}:

$$\frac{1}{\sigma^2}\sum_{i=1}^{n} X_i - \frac{1}{\sigma^2}\sum_{i=1}^{n}\mu = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i = n\mu \quad \Rightarrow \quad \mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

unbiased

Next, solve for σ_{MLE}:

$$\frac{1}{\sigma^3}\sum_{i=1}^{n}(X_i-\mu)^2 = \frac{n}{\sigma} \quad \Rightarrow \sum_{i=1}^{n}(X_i-\mu)^2 = \sigma^2 n \quad \Rightarrow \sigma^2_{MLE} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\mu_{MLE})^2$$

biased

MLE: Bernoulli

Maximum Likelihood Estimate: Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

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

1. Determine formula for $LL(\theta)$
2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0
3. Solve resulting equations

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p)$$

?

$$f(X_i|p) = \begin{cases} p & \text{if } X_i = 1 \\ 1-p & \text{if } X_i = 0 \end{cases}$$

It's difficult to multiply lots of these together, and it's difficult to take logarithms of many of these and add them together.

Ideally, we have a **single** formula that works for both $X_i = 0$ and $X_i = 1$.

Maximum Likelihood Estimate: Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p)$$

$$f(X_i|p) = \begin{cases} p & \text{if } X_i = 1 \\ 1-p & \text{if } X_i = 0 \end{cases}$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$f(X_i|p) = p^{X_i}(1-p)^{1-X_i} \text{ where } X_i \in \{0, 1\}$$

3. Solve resulting equations

✅
- differentiable with respect to p
- valid PMF for the values of X_i we care about

Maximum Likelihood Estimate: Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} \log f(X_i|p) = \sum_{i=1}^{n} \log p^{X_i}(1-p)^{1-X_i}$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$= \sum_{i=1}^{n} [X_i \log p + (1-X_i)\log(1-p)]$$

3. Solve resulting equations

$$= Y(\log p) + (n-Y)\log(1-p), \text{ where } Y = \sum_{i=1}^{n} X_i$$

Maximum Likelihood Estimate: Bernoulli

Consider a sample of n iid RVs $X_1, X_2, \ldots, X_n$.

What is $\theta_{MLE} = p_{MLE}$?

- Let $X_i \sim \text{Ber}(p)$.
- $f(X_i|p) = p^{X_i}(1-p)^{1-X_i}$

1. Determine formula for $LL(\theta)$

$$LL(\theta) = \sum_{i=1}^{n} [X_i \log p + (1 - X_i) \log(1-p)]$$

$$= Y(\log p) + (n - Y) \log(1-p), \text{ where } Y = \sum_{i=1}^{n} X_i$$

2. Differentiate $LL(\theta)$ wrt (each) θ, set to 0

$$\frac{\partial LL(\theta)}{\partial p} = Y\frac{1}{p} + (n - Y)\frac{-1}{1-p} = 0$$

3. Solve resulting equations

$$p_{MLE} = \frac{1}{n}Y = \frac{1}{n}\sum_{i=1}^{n} X_i$$

MLE of the Bernoulli parameter, p_{MLE}, is, once again, the sample mean, $\bar{X}$.

Maximum Likelihood Estimate: Bernoulli

You draw n iid random variables $X_1, X_2, \ldots, X_n$ from some distribution F, yielding the following sample:

$$[1, 0, 1, 1, 1, 1, 0, 1, 1, 1] \qquad (n = 10)$$

Suppose distribution $F = \text{Ber}(p)$ with unknown parameter p.

1. What is p_{MLE}, the MLE of the parameter p?

A. 1.0
B. 0.5
C. 0.8
D. 0.2
E. None/other

$$p_{MLE} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

Maximum Likelihood Estimate: Bernoulli

You draw n iid random variables $X_1, X_2, \ldots, X_n$ from the distribution F, yielding the following sample:

$$[1, 0, 1, 1, 1, 1, 0, 1, 1, 1] \qquad (n = 10)$$

Suppose distribution $F = \text{Ber}(p)$ with unknown parameter p.

1. What is p_{MLE}, the MLE of the parameter p? C. 0.8
2. What is the likelihood $L(\theta)$ of this specific sample?

$$f(X_i|p) = p^{X_i}(1-p)^{1-X_i} \text{ where } X_i \in \{0,1\}$$

$$L(\theta) = \prod_{i=1}^{n} f(X_i|p) \quad \text{where } \theta = p$$

$$= p^8(1-p)^2 = 0.8^8 0.2^2 = 0.00671$$

```
>>> for p in [0.78, 0.79, 0.80, 0.81, 0.82]:
...     print(f'L({p}) = {p ** 8 * (1 - p) ** 2}')
...
L(0.78) = 0.00663
L(0.79) = 0.00669
L(0.80) = 0.00671
L(0.81) = 0.00669
L(0.82) = 0.00662
>>>
```

MLE: Uniform

Maximum Likelihood Estimate: Uniform

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

Let $X_i \sim \text{Uni}(\alpha, \beta)$.

$$f(X_i|\alpha,\beta) = \begin{cases} \frac{1}{\beta - \alpha} & \text{if } \alpha \le x_i \le \beta \\ 0 & \text{otherwise} \end{cases}$$

1. Determine formula for $L(\theta)$

$$L(\theta) = \begin{cases} \left(\frac{1}{\beta - \alpha}\right)^n & \text{if } \alpha \le x_1, x_2, \ldots, x_n \le \beta \\ 0 & \text{otherwise} \end{cases}$$

2. Differentiate $L(\theta)$ wrt each θ, set to 0

A. Great, let's do it
B. Use $LL(\theta)$ instead
C. Constraint $\alpha \le x_1, x_2, \ldots, x_n \le \beta$ makes differentiation hard

Maximum Likelihood Estimate: Uniform

Consider a sample of n iid random variables $X_1, \ldots, X_n$

Assume $X_i \sim \text{Uni}(\alpha, \beta)$.

$$L(\theta) = \begin{cases} \left(\frac{1}{\beta - \alpha}\right)^n & \text{if } \alpha \leq x_1, x_2, \ldots, x_n \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

You observe data: $[0.15, 0.20, 0.30, 0.40, 0.65, 0.70, 0.75]$

Which parameters maximize $L(\theta)$?

A. $\text{Uni}(\alpha = 0.00, \beta = 1.00)$ $\quad (1)^7 = 1$

B. $\text{Uni}(\alpha = 0.15, \beta = 0.75)$ $\quad \left(\frac{1}{0.6}\right)^7 = 59.5$

C. $\text{Uni}(\alpha = 0.15, \beta = 0.70)$ $\quad \left(\frac{1}{0.55}\right)^6 \cdot 0 = 0$

Original, underlying parameters may not yield maximum likelihood.

Maximum Likelihood Estimate: Uniform

Consider a sample of n iid random variables $X_1, X_2, \ldots, X_n$.

Let $X_i \sim \text{Uni}(\alpha, \beta)$.

$$L(\theta) = \begin{cases} \left(\frac{1}{\beta - \alpha}\right)^n & \text{if } \alpha \leq x_1, x_2, \ldots, x_n \leq \beta \\ 0 & \text{otherwise} \end{cases}$$

θ_{MLE}: $\alpha_{MLE} = \min(x_1, x_2, \ldots, x_n)$ $\quad$ $\beta_{MLE} = \max(x_1, x_2, \ldots, x_n)$

Intuition:

- We want interval size $\beta - \alpha$ to be as narrow as possible to maximize the likelihood.
- Need to ensure all datapoints are included in interval. Otherwise, $L(\theta) = 0$.

Maximum Likelihood Estimate: Redux

Maximum Likelihood Estimator θ_{MLE}

- best explains the data we've already seen
- doesn't care at all about any future data

$$\theta_{MLE} = \arg\max_{\theta} L(\theta)$$

✅ In many cases, $\mu_{MLE} = \frac{1}{n}\sum_{i=1}^{n} X_i$ Sample mean (Bernoulli p, Poisson λ, Normal μ)

For some cases, like Uniform: $\alpha_{MLE} \geq \alpha$, $\beta_{MLE} \leq \beta$

⚠️

- Ad hoc, biased, and problematic for small sample sizes
- For example: if $n = 1$, then $\alpha = \beta$, and our estimates yield an invalid distribution

Exercises

Maximum Likelihood Estimate: Solar Panels

We stress test 300 solar panels in a high-temperature aging chamber to imitate the harsh desert conditions where they're often installed.

- Each is monitored until its power output degrades below warranty threshold.
- We collect the failure times (in hours) for all 300 and capture them in a list:

```
lifetimes = [777.09, 1443.87, 1096.06, ..., 1262.27]
```

- Time to failure is well-modeled by an order-3 Weibull distribution* with PDF of

$$f(t_i|\lambda) = \frac{3t_i^2}{\lambda^3} e^{-(t_i/\lambda)^3}$$

This is like an Exponential, except the rate essentially increases over time to reflect age, fatigue, or general wear and tear.

What value of λ maximizes the likelihood of the observed times?

* Its official name is Weibull distribution of shape parameter 3.

Maximum Likelihood Estimate: Solar Panels

Time to failure is well-modeled by an order-3 Weibull distribution with PDF of

$$f(t_i|\lambda) = \frac{3t_i^2}{\lambda^3} e^{-(t_i/\lambda)^3}$$

We measure 300 failure times. What value of λ maximizes their collective likelihood?

1. Define $LL(\lambda)$

$$LL(\lambda) = \sum_{i=1}^{n} \log f(t_i|\lambda) = \sum_{i=1}^{n} \log \frac{3t_i^2}{\lambda^3} e^{-(t_i/\lambda)^3}$$

$$= n \log 3 - 3n \log \lambda + 2 \sum_{i=1}^{n} \log t_i - \frac{1}{\lambda^3} \sum_{i=1}^{n} t_i^3$$

2. Differentiate $LL(\lambda)$ and set to 0

$$\frac{\partial LL(\lambda)}{\partial \lambda} = -\frac{3n}{\lambda} + \frac{3}{\lambda^4} \sum_{i=0}^{n} t_i^3 = 0$$

```
def compute_lambda(lifetimes):
    s = sum([t ** 3 for t in lifetimes])
    return (s / len(lifetimes)) ** (1 / 3)
```

3. Solve $-\frac{3n}{\lambda} + \frac{3}{\lambda^4} \sum_{i=1}^{n} t_i^3 = 0$

multiply by $-\lambda^4$

$$3n\lambda^3 - 3 \sum_{i=1}^{n} t_i^3 = 0$$

$$\lambda^3 = \frac{1}{n} \sum_{i=1}^{n} t_i^3$$

$$\hat{\lambda}_{MLE} = \left(\frac{1}{n} \sum_{i=1}^{n} t_i^3 \right)^{1/3}$$

Maximum Likelihood Estimate: Predicting Rainfall

A Kaua'i-based meteorologist has tracked daily rainfall amounts at the base of Mount Wai'ale'ale for some three years—specifically, 1100 days.

- Daily totals reflect the accumulation of rain from short but frequent rain bursts.
- Daily rainfall amount (in millimeters) is modeled as a random variable whose PDF is

$$f(x_i|\beta) = \frac{\beta^4 x_i^3}{6} e^{-\beta x_i}$$

This models the sum of 4 independent Exponentials, each with rate β. It's a specific instance of a random variable called the Gamma.

What value of β maximizes the likelihood of observing the 1100 rainfall amounts?

Maximum Likelihood Estimate: Predicting Rainfall

Daily rainfall amount (in millimeters) is modeled as a random variable whose PDF is

$$f(x_i|\beta) = \frac{\beta^4 x_i^3}{6} e^{-\beta x_i}$$

We measure total rainfall 1100 days in a row. What β maximizes their likelihood?

1. Define $LL(\beta)$

$$LL(\beta) = \sum_{i=1}^{n} \log f(x_i|\beta) = \sum_{i=1}^{n} \log \frac{\beta^4 x_i^3}{6} e^{-\beta x_i}$$
$$= 4n \log \beta - \beta \sum_{i=0}^{n} x_i + \text{constants}$$

2. Differentiate $LL(\beta)$ and set to 0

$$\frac{\partial LL(\beta)}{\partial \beta} = \frac{4n}{\beta} - \sum_{i=0}^{n} x_i = 0$$

3. Solve

$$\frac{4n}{\beta} - \sum_{i=0}^{n} x_i = 0$$
$$\frac{4n}{\beta} = \sum_{i=0}^{n} x_i$$
$$\hat{\beta}_{MLE} = \frac{4n}{\sum_{i=0}^{n} x_i} = \frac{4}{\bar{X}}$$

Maximum Likelihood Estimation: What's the Big Idea?

The Situation

You have a dataset. You believe this data emerged from a certain type of process or model (like a normal distribution), but you don't know the exact parameters that govern that process.

The Core Philosophy

Maximum Likelihood Estimation works like a detective examining evidence. It asks: **"Of all possible explanations, which one makes my observed data the most believable outcome?"**

01

Observe the Evidence

Flip a coin 10 times and observe 7 heads

02

Consider Explanations

Could be fair coin (p=0.5) or biased coin (p=0.7)

03

Choose Most Likely

MLE selects p=0.7 as it makes your data least surprising

MLE finds parameter values that make your observed data the least surprising outcome possible.

Why Statisticians Trust MLE

MLE isn't just a clever idea—its estimates possess mathematically proven desirable properties, particularly with large datasets.

Consistency

The more data you provide, the smarter it becomes. With sufficient data, MLE estimates converge to the true parameter value, unswayed by random noise.

Efficiency

Among all unbiased estimators, MLE achieves the smallest variance. It extracts maximum information from your data—no information is wasted.

Asymptotic Normality

With large samples, MLE estimates follow a predictable normal distribution, enabling precise confidence interval calculations.

Invariance

If you find the MLE for one parameter, you automatically obtain the MLE for any function of that parameter—no additional computation required.

These properties make MLE the gold standard for parameter estimation in modern statistics and machine learning.

MLE in Practice: Strengths & Limitations

The Universal Toolkit

MLE's greatest strength lies in its **unified philosophical approach**. This single powerful framework scales from simple textbook problems to training complex neural networks.

- Works across diverse probability distributions
- Foundation for modern machine learning algorithms
- Consistent methodology regardless of problem complexity

The Critical Assumption

MLE completely trusts your model specification. If you assume normality when data follows a different distribution, MLE will still find the "optimal" normal parameters—but the results will be misleading.

> *"MLE gives you the best possible answer, assuming you asked the right question in the first place."*

Model Selection Matters

Success depends entirely on choosing appropriate probability models that reflect your data's true underlying structure.

Diagnostic Checking Essential

Always validate model assumptions through residual analysis, goodness-of-fit tests, and visual diagnostics before trusting MLE results.

Applications of Maximum Likelihood Estimation Across Disciplines

Maximum Likelihood Estimation extends far beyond theoretical statistics, serving as the cornerstone for parameter estimation across numerous quantitative disciplines. From financial risk models to medical trials, MLE provides the mathematical foundation for evidence-based decision making.

Finance & Econometrics

Risk Management & Asset Pricing: Financial models assume asset returns follow specific distributions (e.g., log-normal). MLE estimates critical parameters—expected return and volatility—from historical data, feeding into Value at Risk models and derivative pricing frameworks.

Reliability Engineering

Survival Analysis: Models "time-to-event" data using Weibull or Exponential distributions. MLE estimates parameters for predicting component failures and evaluating medical treatment efficacy in clinical trials.

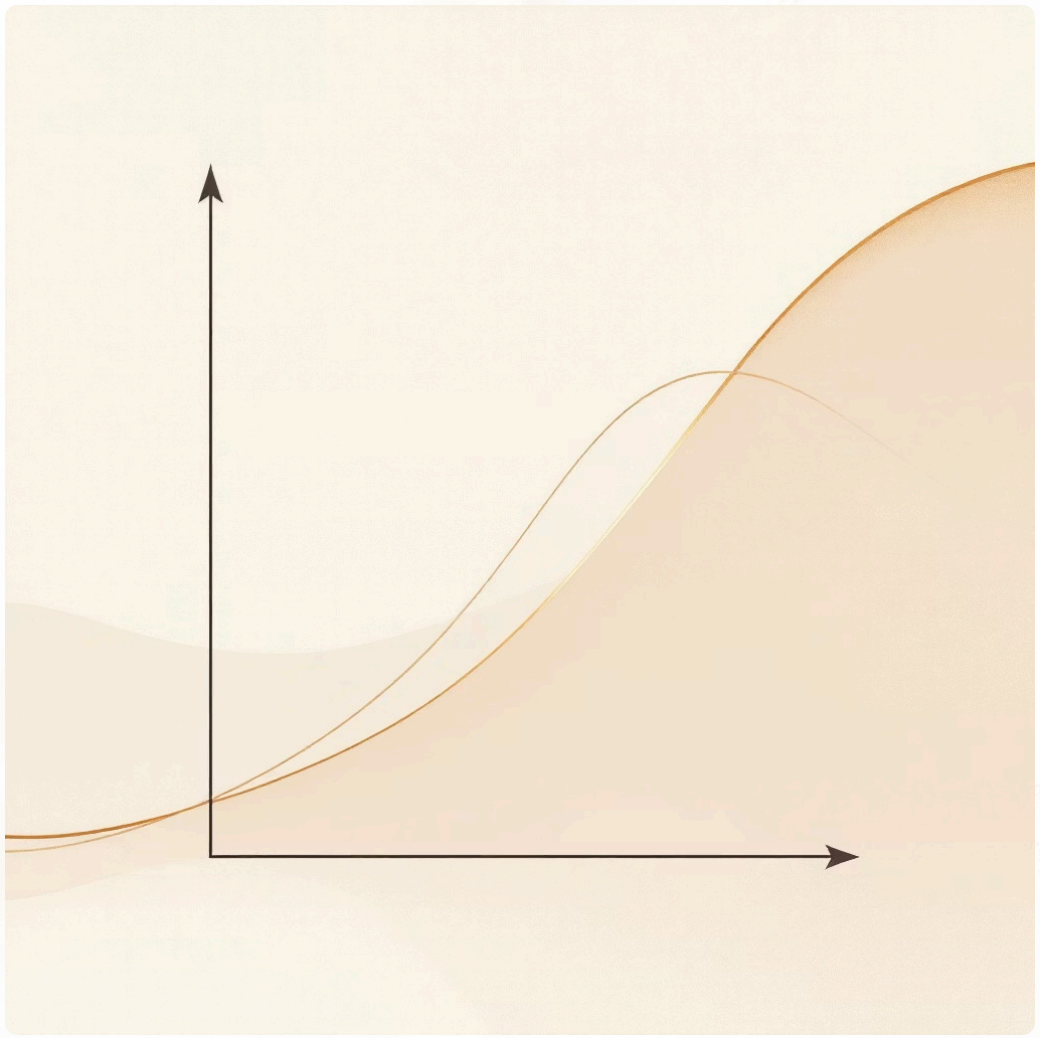
Signal Processing

Communications Theory: In digital communications, transmitted signals are corrupted by Gaussian noise. MLE decodes the most likely original signal from noisy observations, ensuring accurate data transmission.

MLE as the Foundation of Machine Learning

A profound connection exists between likelihood maximisation and supervised learning. Many standard loss functions derive directly from MLE principles, revealing the statistical underpinnings of modern algorithms.

Logistic Regression



Binary Classification: Training logistic regression minimises Binary Cross-Entropy loss, which is mathematically equivalent to maximising the joint likelihood of observed class labels.

The resulting model coefficients are Maximum Likelihood Estimates for logistic model parameters.

Linear Regression

Continuous Prediction: Ordinary Least Squares minimises Sum of Squared Errors. Under Gaussian error assumptions, OLS solutions are identical to Maximum Likelihood Estimates.

This provides probabilistic justification for the least squares criterion's widespread adoption.

Minimising a model's loss function in machine learning often represents a direct application of Maximum Likelihood Estimation principles.

Contributions

Group 5

2201AI03, Amit Vinod	: 4.16%
2201AI05, Anand Kumar	: 4.16%
2201AI10, Divyanshu Gupta	: 4.16%
2201AI12, Hari Om Kumar	: 4.16%
2201AI26, Nirjay Kumar	: 4.16%
2201AI32, Rishabh Verma	: 4.16%
2201AI57, Arunangshu Pal	: 4.16%
2201CS46, Md Adil Ansari	: 4.16%
2201CS47, Md Kamran	: 4.16%
2201CS72, Swarup Suthar	: 4.16%
2201CS74, Umar Khan	: 4.16%
2201CS87, Uday Shrotriya	: 4.16%

Group 6

2201AI13, Harpranav	: 4.16%
2201AI22, Lokesh	: 4.16%
2201CS23, Divyam	: 4.16%
2201CS24, Erum	: 4.16%
2201CS33, Jatin	: 4.16%
2201CS36, Kashish	: 4.16%
2201CS48, Faizan	: 4.16%
2201CS49, Moulik	: 4.16%
2201CS50, Nagesh	: 4.16%
2201CS55, Pranjal	: 4.16%
2201CS56, Rachit	: 4.16%
2201CS80, Yuvraj	: 4.16%