20: Maximum Likelihood Estimation

Jerry Cain
February 26, 2024

Lecture Discussion on Ed
Parameter Estimation
Story so far

At this point:
If we’re provided with a **model** and all the necessary parameters, we can make predictions!

\[ Y \sim \text{Exp} \left( \frac{1}{5} \right) \]
\[ X_1, \ldots, X_n \text{ iid} \]
\[ X_i \sim \text{Ber}(0.2), \]
\[ X = \sum_{i=1}^{n} X_i \]

But how do we **infer** and **estimate** the parameters for a given model?

What if you want to learn the **structure** of the model, too?

---

Machine Learning

Lisa Yan, Chris Piech, Mehran Sahami, and Jerry Cain, CS109, Winter 2024
Some estimators

$X_1, X_2, \ldots, X_n$ are $n$ iid random variables, where $X_i$ drawn from distribution $F$ with $E[X_i] = \mu$, $\text{Var}(X_i) = \sigma^2$.

Sample mean:

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

unbiased estimate of $\mu$

Sample variance:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

unbiased estimate of $\sigma^2$

potentially useful estimates if trying to infer parameters of a Gaussian
What are parameters?

def Most random variables we’ve seen thus far are \textbf{parametric models}:

\[ \text{Distribution} = \text{model} + \text{parameter } \theta \]

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

For each of the distributions below, what is the parameter $\theta$?

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

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

\[
\text{Distribution} = \text{model} + \text{parameter } \theta
\]

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

For each of the distributions below, what is the parameter \( \theta \)?

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

\( \theta \) is the parameter of a distribution. \( \theta \) can be a vector.
Why do we care?

In the real world, we don’t know the true parameters.
• But we **observe data**: # times coin comes up heads, # requests for RydeShare per minute, # visitors to website per day, offer amount for that used bike you can’t sell

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

In parameter estimation,
We’ll initially and often rely on **point estimates**—i.e., the best single value
• Provides an understanding of why data looks the way it does
• Can make future **predictions** using that model
• Can run simulations to generate more data
Maximum Likelihood Estimator
Defining the likelihood of data: 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 \( \theta \).
- Observed sample:

\[
[0, 0, 1, 1, 1, 1, 1, 1, 1, 1] \quad (n = 10)
\]

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

\[
P(\text{sample}|\theta = 0.4) = (0.4)^8 (0.6)^2 = 0.000236
\]

Is there a better choice for \( \theta \)?

![Python code](Python code)

**Likelihood of data given parameter \( \theta = 0.4 \)**
Defining the likelihood of data

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: $(X_1, X_2, \ldots, X_n)$

Likelihood question:
How likely is the sample $(X_1, X_2, \ldots, X_n)$ given the parameter $\theta$?

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.
Maximum Likelihood Estimator

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

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

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

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

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

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

Likelihood of your sample

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

For continuous $X_i$, $f(X_i|\theta)$ is PDF, and for discrete $X_i$, $f(X_i|\theta)$ is PMF.
Maximum Likelihood Estimator

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

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

\[
\theta_{MLE} = \arg \max_{\theta} L(\theta)
\]
argmax and log likelihood
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. \( \arg \max_x f(x) \)?
   \[ = 0 \]
Argmax properties

\[
\arg \max_x f(x) \quad \text{The argument } x \text{ that maximizes the function } f(x).
\]

\[
= \arg \max_x \log f(x) \quad \text{(log is an increasing function: } x < y \iff \log x < \log y)\\
\]

\[
= \arg \max_x (c \log f(x)) \quad \text{(} x < y \iff c \log x < c \log y)\\
\]

for any positive constant \(c\)
Finding the argmax with calculus

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

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

Differentiate w.r.t. argmax’s argument

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

Set to 0 and solve

$$2x = 0 \implies \hat{x} = 0$$

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

- Check $f(\hat{x} \pm \epsilon) < f(\hat{x})$
- Often ignored in expository derivations
- We’ll ignore it as well (and we won’t require it in class or on problem sets and exams)
Maximum Likelihood Estimator

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

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

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

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

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

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

Stanford University
MLE: Bernoulli
Computing the MLE

General approach for finding $\theta_{MLE}$, the MLE of $\theta$:

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

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

2. Differentiate $LL(\theta)$ w.r.t. (each) $\theta$

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

To maximize:

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

3. Solve resulting equations

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

$LL(\theta)$ is often easier to differentiate than $L(\theta)$. 
Maximum Likelihood with Bernoulli

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

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

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

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

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

3. Solve resulting equations

Let $X_i \sim Ber(p)$.

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

Consider a sample of $n$ iid RVs $X_1, X_2, \ldots, X_n$. What is $\theta_{MLE} = p_{MLE}$?

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

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

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

3. Solve resulting equations

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

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

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

- differentiable with respect to $p$
- valid PMF over discrete domain
Maximum Likelihood with Bernoulli

Consider a sample of \( n \) iid RVs \( X_1, X_2, \ldots, X_n \). What is \( \theta_{MLE} = p_{MLE} \)?

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) \( \theta \), set to 0

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

3. Solve resulting equations

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

- Let \( X_i \sim \text{Ber}(p) \).
- \( f(X_i|p) = p^{X_i}(1 - p)^{1-X_i} \)
Maximum Likelihood with Bernoulli

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

What is \( \theta_{MLE} = p_{MLE} \)?

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) \( \theta \), set to 0

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

3. Solve resulting equations

   * Let \( X_i \sim \text{Ber}(p) \).
   * \( f(X_i|p) = p^{X_i}(1 - p)^{1-X_i} \)
Maximum Likelihood with Bernoulli

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

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

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

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

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

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

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

3. Solve resulting equations

   • Let $X_i \sim \text{Ber}(p)$.
   • $f(X_i|p) = p^{X_i} (1 - p)^{1-X_i}$
Maximum Likelihood with Bernoulli

Consider a sample of $n$ iid RVs $X_1, X_2, ..., X_n$. 
What is $\theta_{MLE} = p_{MLE}$?

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

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

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

2. Differentiate $LL(\theta)$ wrt (each) $\theta$, 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 the sample mean, $\bar{X}$, which is an unbiased estimator of the true mean.
Quick check

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

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

• 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

\[ p_{MLE} = \bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]
Quick check

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

\[
[0, 0, 1, 1, 1, 1, 1, 1, 1, 1] \quad (n = 10)
\]

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

1. What is \( p_{\text{MLE}} \), the MLE of the parameter \( p \)?
   
   C. 0.8

2. What is the likelihood \( L(\theta) \) of this specific sample?

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

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

\[
= p^8(1 - p)^2 = 0.8^80.2^2 = 0.0067
\]
MLE: Poisson and Uniform
Maximum Likelihood with Poisson

Consider a sample of \( n \) iid RVs \( X_1, X_2, \ldots, X_n \). What is \( \theta_{MLE} = \lambda_{MLE} \)?

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

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

using natural log,

i.e., \( \ln e = 1 \)

- Let \( X_i \sim \text{Poi}(\lambda) \).
- PMF: \( f(X_i | \lambda) = \frac{e^{-\lambda} \lambda^{X_i}}{X_i!} \)
Maximum Likelihood with Poisson

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

What is \( \theta_{MLE} = \lambda_{MLE} \)?

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

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

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

using natural log, i.e., \( \ln e = 1 \)

2. Differentiate \( LL(\theta) \) w.r.t. (each) \( \theta \), 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 with Poisson

Consider a sample of $n$ iid RVs $X_1, X_2, \ldots, X_n$. What is $\theta_{MLE} = \lambda_{MLE}$?

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

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

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

   Using natural log, i.e., $\ln e = 1$

2. Differentiate $LL(\theta)$ w.r.t. (each) $\theta$, set to 0

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

3. Solve resulting equations

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

   MLE of the Poisson parameter, $\lambda_{MLE}$, is the sample mean, $\bar{X}$, which is an unbiased estimator of the true mean.
Quick Review

1. A particular experiment can be modeled as a Poisson RV with parameter $\lambda$, in terms of events/minute. Collect data: observe 53 events over the next 10 minutes. What is $\lambda_{MLE}$?

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

2. Is the Bernoulli MLE an unbiased estimator of the Bernoulli parameter $p$? ✓

3. Is the Poisson MLE an unbiased estimator of the Poisson variance? ✓

4. What does unbiased mean?

$$E[\text{estimator}] = \text{the truth}$$

Unbiased: If you repeat your experiment multiple times, on average, you’ll get what you are looking for.
Maximum Likelihood with Uniform

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

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

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

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

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

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

A. Great, let’s do it
B. Use $LL(\theta)$ instead
C. Constraint $\alpha \leq x_1, x_2, ..., x_n \leq \beta$ makes differentiation hard
Maximum Likelihood with Uniform: Sample

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

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

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

Underlying $X_i \sim \text{Uni}(0,1)$

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

Which parameters maximize $L(\theta)$?

A. $\text{Uni}(\alpha = 0.00, \beta = 1.00)$
B. $\text{Uni}(\alpha = 0.15, \beta = 0.75)$
C. $\text{Uni}(\alpha = 0.15, \beta = 0.70)$
Maximum Likelihood with Uniform: Sample

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

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

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

Underlying $X_i \sim \text{Uni}(0,1)$

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

Which parameters maximize $L(\theta)$?

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

$$L(\theta) = \left( \frac{1}{1 - 0.00} \right)^7 = 1$$

B. $\text{Uni}(\alpha = 0.15, \beta = 0.75)$

$$L(\theta) = \left( \frac{1}{0.75 - 0.15} \right)^7 = 59.5$$

C. $\text{Uni}(\alpha = 0.15, \beta = 0.70)$

$$L(\theta) = \left( \frac{1}{0.55} \right)^6 \cdot 0 = 0$$

⚠ Original parameters may not yield maximum likelihood.
Maximum Likelihood with Uniform

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

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

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

$\theta_{\text{MLE}}$: $\alpha_{\text{MLE}} = \min(x_1, x_2, ..., x_n) \quad \beta_{\text{MLE}} = \max(x_1, x_2, ..., x_n)$

Intuition:

• Want interval size $\beta - \alpha$ to be as narrow as possible to maximize likelihood function.

• Need to ensure all datapoints are included in interval. Otherwise, $L(\theta) = 0$. (demo)
Small samples = problems with MLE

Maximum Likelihood Estimator \( \theta_{MLE} \):
- Best explains the data we’ve seen
- Does not attempt to generalize to data not yet observed.

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

In many cases,

\[
\mu_{MLE} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

Sample mean

(MLE for Bernoulli \( p \), Poisson \( \lambda \), Normal \( \mu \))

- Unbiased (\( E[\mu_{MLE}] = \mu \), regardless of sample size)

For some cases, like Uniform:

\[
\alpha_{MLE} \geq \alpha, \quad \beta_{MLE} \leq \beta
\]

- Ad hoc, biased, and problematic for small sample sizes
- Example: If \( n = 1 \), then \( \alpha = \beta \), and our estimates yield an invalid distribution
MLE: Gaussian
Maximum Likelihood with Normal

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

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

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

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

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

(assuming natural log)

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

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

3. Solve resulting equations
Maximum Likelihood with Normal

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

- 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 $LL(\theta)$

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

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

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

3. Solve resulting equations

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

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

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

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

1. Determine formula for \( LL(\theta) \) with respect to \( \mu \)
   \[
   LL(\theta) = -\sum_{i=1}^{n} \log(\sqrt{2\pi\sigma}) - \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{2\sigma^2}
   \]
   \[
   \frac{\partial LL(\theta)}{\partial \mu} = \sum_{i=1}^{n} \frac{2(X_i - \mu)}{2\sigma^2} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0
   \]

2. Differentiate \( LL(\theta) \) wrt (each) \( \theta \), set to 0
   \[
   \frac{\partial LL(\theta)}{\partial \sigma} = -\sum_{i=1}^{n} \frac{1}{\sigma} + \sum_{i=1}^{n} \frac{2(X_i - \mu)^2}{2\sigma^3}
   \]
   \[
   = -\frac{n}{\sigma} + \frac{1}{\sigma^3} \sum_{i=1}^{n} (X_i - \mu)^2 = 0
   \]

3. Solve resulting equations with respect to \( \sigma \)
Maximum Likelihood with Normal

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

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

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

3. Solve resulting equations

Two equations, two unknowns:

\[
\frac{1}{\sigma^2} \sum_{i=1}^{n} (X_i - \mu) = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i = n\mu
\]

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

First, solve for \( \mu_{\text{MLE}} \):

\[
\frac{1}{\sigma^2} \sum_{i=1}^{n} X_i - \frac{1}{\sigma^2} \sum_{i=1}^{n} \mu = 0 \quad \Rightarrow \quad \sum_{i=1}^{n} X_i = n\mu
\]

\( \mu_{\text{MLE}} = \frac{1}{n} \sum_{i=1}^{n} X_i \) unbiased
Maximum Likelihood with Normal

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

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

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

3. Solve resulting equations

Two equations, two unknowns:

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

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

First, solve for \( \mu_{MLE} \):

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

unbiased

Next, solve for \( \sigma^2_{MLE} \):

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

biased