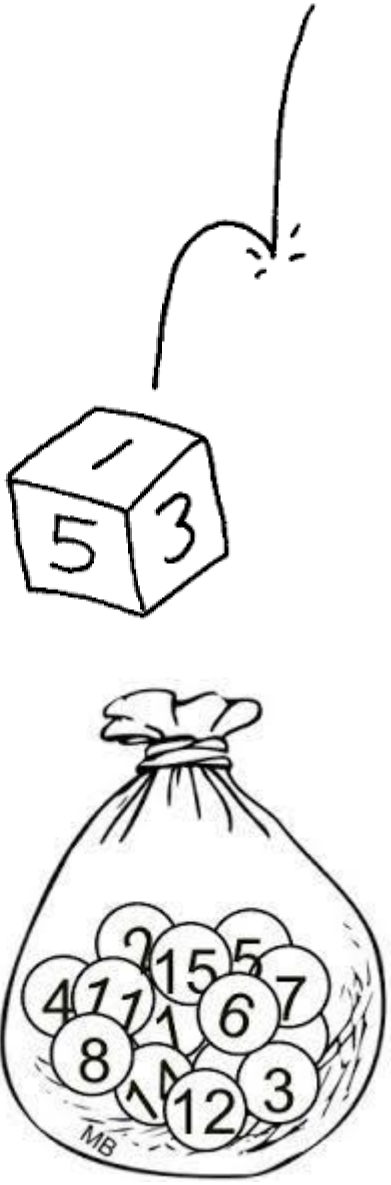
Lecture 16

- Parameter Estimation
- Maximum Likelihood Estimator



Parameter Estimation

Data → Parameters

Parameter estimation

- Point estimation 点估计
  - Moment Estimator
  - Maximum Likelihood Estimator
  - other
- Interval estimation 区间估计

Recall some estimators

$X_1, X_2, \ldots, X_n$ are i.i.d. R.V.s.

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

unbiased estimate of μ

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

unbiased estimate of σ^2

What are parameters?

Def. Many random variables we have learned so far are parametric models:

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

Ex. The distribution $X \sim b(0.2)$ = Bernoulli model + parameter $\theta = 0.2$

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

1. $\mathrm{Ber}(p)$ $\quad\theta = p$
2. $\pi(\lambda)$ $\quad\theta = \lambda$
3. $U(a, b)$ $\quad\theta = (a, b)$
4. $\mathcal{N}(\mu, \sigma^2)$ $\quad\theta = (\mu, \sigma^2)$
5. $Y = mX + b$ $\quad\theta = (m, b)$

θ is the parameter of a distribution.

θ can be a vector of parameters!

Why do we care?

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

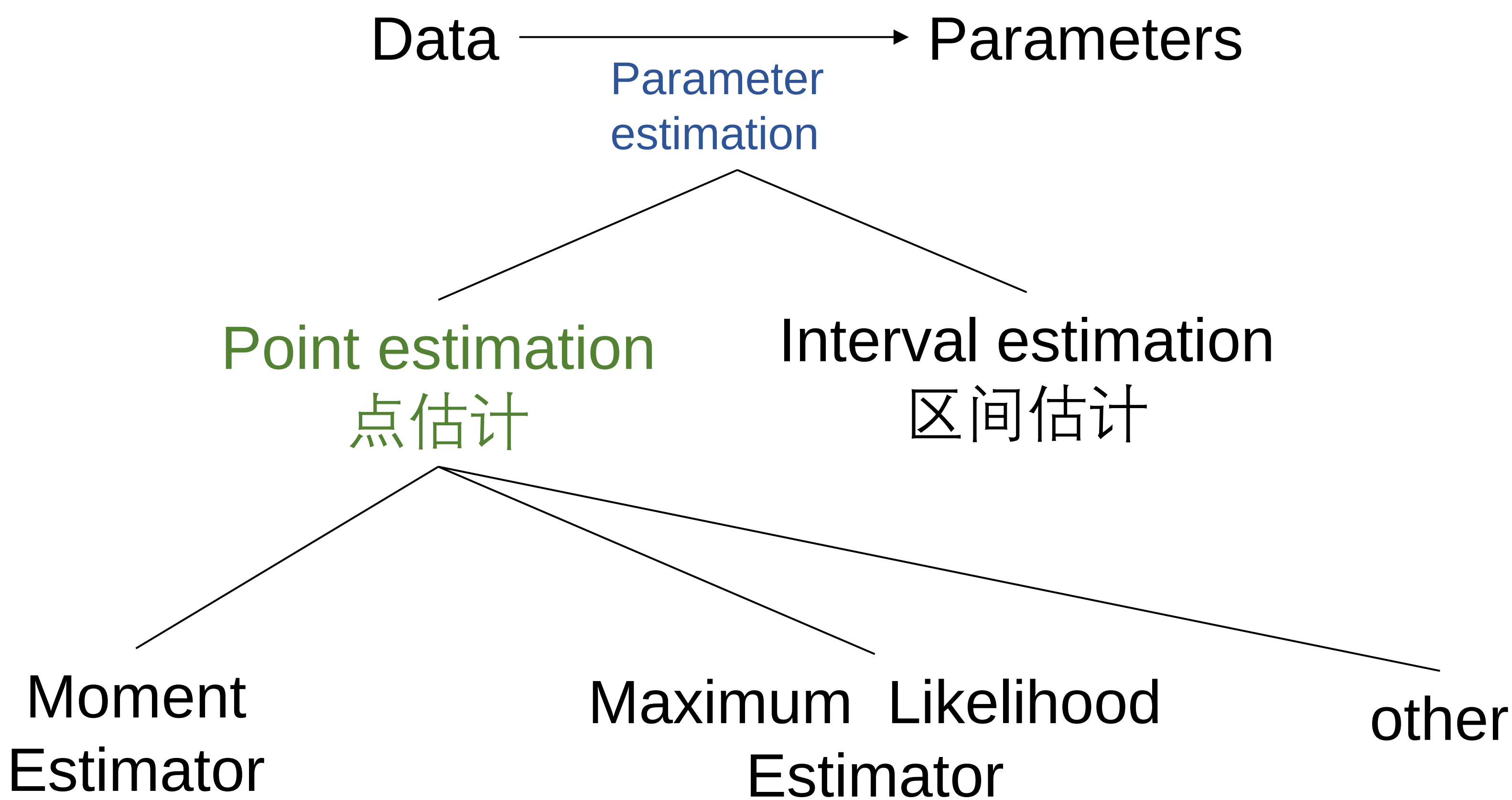
- But observe data: (# times coin comes up heads, lifetimes of disk drives produced, # visitors to website per day, etc.)

def estimator $\hat{\theta}$: random variable estimating parameter θ from data.

In parameter estimation,
point estimate (点估计) ⇒ estimated parameter:
(the best single value)

- Better understanding of the process producing data
- Future predictions based on model
- Simulation of future processes

Parameter Estimation



the state or fact of something's being likely; probability.

------- Oxford Languages

Maximum Likelihood Estimator

极大似然估计

Defining the likelihood of data: Bernoulli

Consider a sample of n i.i.d. random variables $X_1, X_2, \dots, X_n$.

- X_i from distribution $F = \mathrm{Ber}(\theta)$ with unknown parameter.
- Observed data:

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

How likely was the observed data if $\theta = 0.4$?

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

Is $\theta = 0.4$ a good parameter? Is there a "**better**" parameter θ?

calculate $P(\text{sample}|\theta)$ for $\theta \in [0,1] \Longrightarrow$ find the "best" θ.

Defining the likelihood of data: Bernoulli

Consider a sample of n i.i.d. random variables $X_1, X_2, \dots, X_n$.

- X_i from a distribution with PDF/PMF $f(X_i|\theta)$.
- Observed data: $(X_1, X_2, \dots, X_n)$

Likelihood question:
How likely was data $(X_1, X_2, \dots, X_n)$ given parameter θ?

That is the "**probability**": $f(X_1, X_2, \dots, X_n|\theta)$

Likelihood function (似然函数), $L(\theta)$:

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

X_i are i.i.d.

Maximum Likelihood Estimator

Consider a sample of n i.i.d. 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 finds:

the value of θ that maximizes $L(\theta)$

$$\theta_{\mathrm{MLE}} = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$$

Likelihood of your sample

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

For continuous X_i, $f(X_i|\theta)$ is PDF;
for discrete X_i, $f(X_i|\theta)$ is PMF.

Maximum Likelihood Estimator

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

<u>def</u> The Maximum Likelihood Estimator (MLE) of θ finds:

the value of θ that maximizes $L(\theta)$

$$\theta_{\text{MLE}} = \underset{\theta}{\operatorname{argmax}} L(\theta)$$

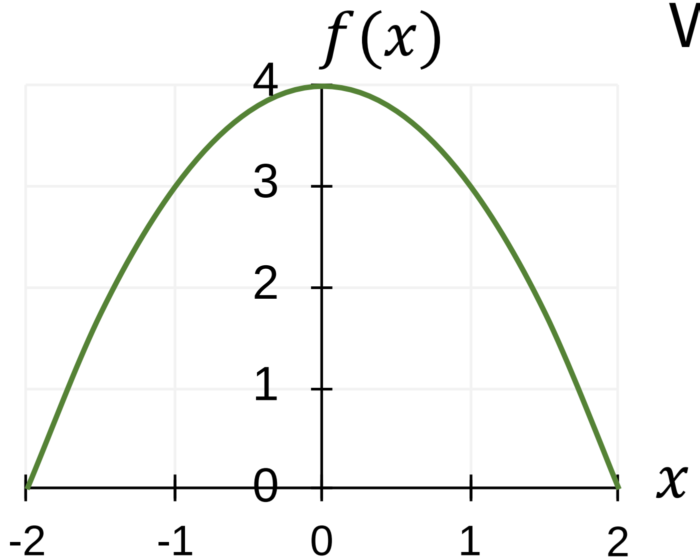
Find the argument $\theta = \theta_{MLE}$ that maximizes $L(\theta)$

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 $-2 < x < 2$



Write down the value of :

$\max_x f(x)$?

=4

$\arg\max_x f(x)$?

=0

Solve argmax with derivative

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

Let $f(x) = -x^2 + 4$
Where $-2 < x < 2$

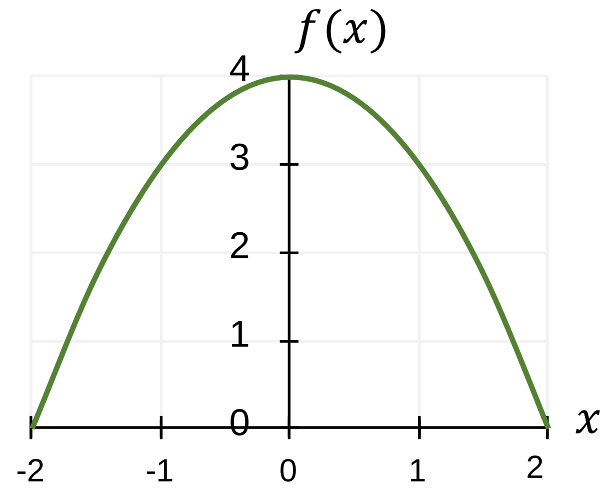
Differentiate w.r.t. argmax's argument $\quad \dfrac{d}{dx} f(x) = \dfrac{d}{dx}(-x^2 + 4) = -2x$

Set to 0 and solve $\quad -2x = 0 \quad \Rightarrow \quad \hat{x} = 0$

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

Check $f(\hat{x} \pm \varepsilon) < f(\hat{x})$, or

$$\frac{d}{dx} f(\hat{x} - \varepsilon) > 0, \frac{d}{dx} f(\hat{x} + \varepsilon) < 0$$



Maximum Likelihood Estimator

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

<u>def</u> The Maximum Likelihood Estimator (MLE) of θ finds:

the value of θ that maximizes $L(\theta)$

$$\theta_{\mathrm{MLE}} = \underset{\theta}{\operatorname{argmax}} L(\theta)$$

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

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

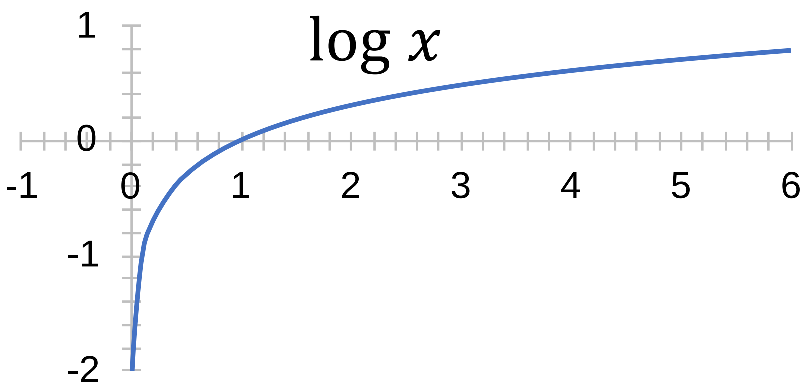
However, taking derivative of $L(\theta)$ is complicated. A simple **transform** will be helpful!

The log function, some properties

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

After taking log, the multiplicative of n terms become a summation of n terms.

Let $\log(\cdot)$ denotes the log function with arbitrary base.



- Log is monotonic:
  $x \leq y \;\Leftrightarrow \log x \leq \log y$
- Log of product = sum of logs:
  $\log(ab) = \log a + \log b$
- Log of exponential function:
  $\log a^b = b \cdot \log a$
- Natural logs:
  $\log_e x = \ln x$

argmax+ln properties

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

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

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

Log is monotonic:
$x \le y \Leftrightarrow \ln x \le \ln y$

$$= \arg\max_x (c \cdot \ln f(x))$$

for any positive constant c

$x \le y \Leftrightarrow c \ln x \le c \ln y$

How do we really compute $\arg\max f(x)$?

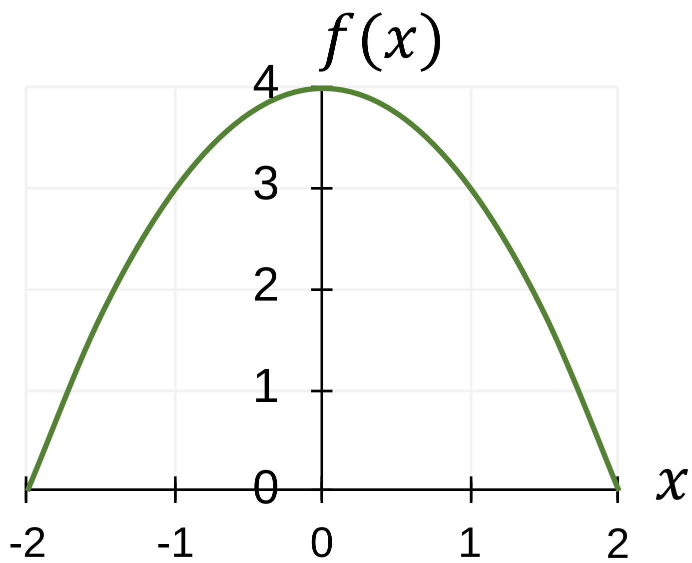
argmax and log

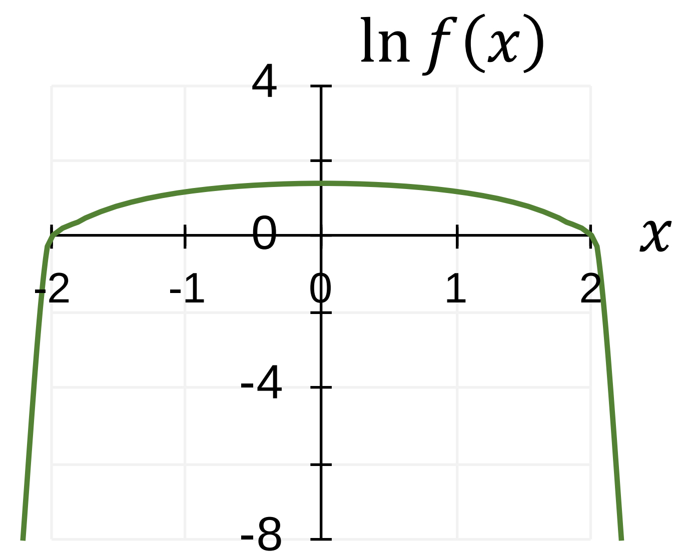
$$\arg\max_x f(x)$$
$$= \arg\max_x \ln f(x)$$

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

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

$$\arg\max_x f(x) = 0$$





Steps in MLE

1. Establish the likelihood function, $L(\theta)$:

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

2. Take natural log of the likelihood function:

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

3. Find the solution with derivative on log likelihood $LL(\theta)$:

$$\frac{d \ln L(\theta)}{d\theta} = 0 \quad \text{or} \quad \frac{\partial \ln L(\theta_i)}{\partial \theta_i} = 0, \ (i = 1,2,\ldots,k)$$

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

Maximum Likelihood with Bernoulli

Consider a sample of n i.i.d. R.V.s $X_1, X_2, \dots, X_n$.
What is θ_{MLE}?

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

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

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

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

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

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

3. Solve resulting equations

Maximum Likelihood with Bernoulli

Consider a sample of n i.i.d. R.V.s $X_1, X_2, \ldots, X_n$.
What is θ_{MLE}?

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

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

$$LL(p) = \sum_{i=1}^{n} \ln f(X_i|p) = \sum_{i}^{n} \ln[p^{X_i}(1-p)^{1-X_i}]$$

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

$$= \sum_{i=1}^{n} X_i (\ln p) + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p)$$

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

3. Solve resulting equations

Maximum Likelihood with Bernoulli

Consider a sample of n i.i.d. R.V.s $X_1, X_2, \ldots, X_n$.
What is θ_{MLE}?

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

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

$$LL(p) = \sum_{i=1}^{n} X_i\,(\ln p) + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p)$$

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

$$\frac{\partial LL(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i + \left(n - \sum_{i=1}^{n} X_i\right)\frac{-1}{1-p} = 0$$

3. Solve resulting equations

Maximum Likelihood with Bernoulli

Consider a sample of n i.i.d. R.V.s $X_1, X_2, \dots, X_n$.
What is θ_{MLE}?

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

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

$$LL(p) = \sum_{i=1}^{n} X_i (\ln p) + \left(n - \sum_{i=1}^{n} X_i\right) \ln(1-p)$$

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

$$\frac{\partial LL(p)}{\partial p} = \frac{1}{p}\sum_{i=1}^{n} X_i + \left(n - \sum_{i=1}^{n} X_i\right)\frac{-1}{1-p} = 0$$

3. Solve resulting equations

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

MLE of the Bernoulli parameter, p_{MLE} is the unbiased estimate of the mean, $\bar{X}$ (sample mean).

Quick check

$p_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} X_i$ for $X \sim \text{Ber}(n, p)$

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

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

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 of the above

$$p_{\text{MLE}} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} = 0.8$$

Quick check

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

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

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

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

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

$$L(p) = \prod_{i=1}^{n} f(X_i|p) = p^8(1-p)^2$$

Ex. Consider drawing a sample of n i.i.d. R.V.s. $X_1, X_2, \ldots, X_n$ from population $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are unknown. Given the observed data $x_1, x_2, \ldots, x_n$, find the MLE of parameters μ and σ^2.

Ex. Consider drawing a sample of n i.i.d. R.V.s. $X_1, X_2, \dots, X_n$ from population $X \sim \mathcal{N}(\mu, \sigma^2)$, where μ and σ^2 are unknown. Given the observed data $x_1, x_2, \dots, x_n$, find the MLE of parameters μ and σ^2.

Sol. The PDF of X is

$$f(x;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},\ \text{where } \theta = (\mu, \sigma^2).$$

The likelihood function is

$$L(\mu,\sigma^2) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

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

$$\begin{cases} \dfrac{\partial}{\partial\mu}\ln L(\mu,\sigma^2) = \dfrac{1}{\sigma^2}\displaystyle\sum_{i=1}^{n}(x_i-\mu) = 0 \\ \dfrac{\partial}{\partial\sigma^2}\ln L(\mu,\sigma^2) = -\dfrac{n}{2\sigma^2} + \dfrac{1}{2\sigma^4}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2 = 0 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X}, \\ \hat{\sigma}^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i-\bar{X})^2 \end{cases}$$

Limitation of MLE

Def. The Maximum Likelihood Estimator (MLE) of θ finds:

the value of θ that maximizes $L(\theta)$

$$\theta_{\mathrm{MLE}} = \underset{\theta}{\operatorname{argmax}}\, L(\theta)$$

Likelihood of your sample

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

Limitations:

1) Require knowledges of $f(X_i|\theta)$!
2) $d\, LL(\theta)/d\theta$ is sometimes difficult to solve.

Any alternative ways?

Try by yourselves

Ex. Given a population follows

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha}, 0 < x < 1 \\ 0, \qquad\qquad\qquad \text{otherwise} \end{cases}$$

where the value of α is unknown. $X_1, X_2, \ldots, X_n$ is a sample from X. Find the MLE estimator of α.

Ex. Given the PMF of X as

X	1	2	3
p_i	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

where θ $(0 < \theta < 1)$ is unknown. Given the observed data

$$x_1 = 1, x_2 = 2, x_3 = 1$$

Find the MLE estimator of θ.

Try by yourselves

<u>Ex</u>. Given a population follows

$$f(x) = \begin{cases} (\alpha+1)x^{\alpha}, 0 < x < 1 \\ 0, \qquad\qquad\qquad \text{otherwise} \end{cases}$$

where the value of α is unknown. $X_1, X_2, \ldots, X_n$ is a sample from X. Find the MLE estimator of α.

Maximum Likelihood Estimator

$L(\theta)$: $\quad L(\alpha) = \prod_{i=1}^{n}(\alpha+1)X_i^{\alpha}$

$$= (\alpha+1)^n \prod_{i=1}^{n} X_i^{\alpha}$$

$LL(\theta)$: $\quad \ln L(\alpha) = n \cdot \ln(\alpha+1) + \alpha \sum_{i=1}^{n} \ln X_i$

Derivative: $\quad \frac{\partial \ln L(\alpha)}{\partial \alpha} = \frac{n}{\alpha+1} + \sum_{i=1}^{n} \ln X_i = 0$

Find $\hat{\theta}$: $\quad \hat{\alpha} = -\frac{n}{\sum_{i=1}^{n} \ln X_i} - 1$

Try by yourselves

Ex. Given the PMF of X as

X	1	2	3
p_i	θ^2	$2\theta(1-\theta)$	$(1-\theta)^2$

where θ $(0 < \theta < 1)$ is unknown. Given the observed data

$$x_1 = 1, x_2 = 2, x_3 = 1$$

Find the MLE estimator of θ.

Maximum Likelihood Estimator:

Likelihood function:

$$L(\theta) = \prod_{i=1}^{3} P\{X_i = x_i\}$$
$$= P\{X_1 = 1\}P\{X_2 = 2\}P\{X_3 = 1\}$$
$$= 2\theta^5(1-\theta).$$

Take natural log:

$$\ln L(\theta) = \ln 2 + 5 \cdot \ln \theta + \ln(1-\theta).$$

Take derivative:

$$\frac{d \ln L(\theta)}{d\theta} = 0 + \frac{5}{\theta} - \frac{1}{1-\theta} = 0, \quad \therefore \ \hat{\theta} = \frac{5}{6}.$$