## © The Method of Maximum Likelihood

Situation 1. Finding the number of fishes in a pond
Q: There are more than 10,000 fish in a pond.
How to estimate the number of fishes if you are the manager of the pond?
Exp.: Step 1. Catch 1000 fishes
Step 2. Mark them and put them back to the pond
Step 3. Catch another 1000 fishes in a few days and see how many have marks on them, say, 10 .
Step 4. Fraction of marked fishes: $10 / 1000=1 / 100$
Step 5. Estimated number of fishes $=1000 /(1 / 100)=100,000$ Is this number likely?

Situation 2. A bag having 4 balls (Black or White)
Experiment: Extracting 3 balls

- One per each trial without putting back into bag

Results: 3 white balls
Hypotheses: 1. Parent: 3 white balls \& 1 black ball 2. All 4 balls: White.

## Which one is more likelihood?

Probability of getting 3 white balls from Hypo.1: $\frac{3}{4} \cdot \frac{2}{3} \cdot \frac{1}{2}=\frac{1}{4}$
Probability of getting 3 white balls from Hypo.2: $\frac{4}{4} \cdot \frac{3}{3} \cdot \frac{2}{2}=1$
(More likelihood)

## - Method of Maximum Likelihood

- Determining which of two or more competing hypotheses (estimations) from experiments yields best fits to the data.


## ※ Bayes’ Theorem

$P(H \mid A)=\frac{P(H) P(A \mid H)}{P(A)}=\frac{P(H) L(H \mid A)}{P(A)}$
where $H$ : Hypothesis \& $A$ : Observation
$P(H)$ : Prior (Marginal) probability of $H$ without information about $A$ $P(H \mid A)$ : Posterior (Conditional) probability of $H$ depending on $A$
$P(A \mid H)$ : Conditional probability of $A$ given $H$
$L(H \mid A)$ : Likelihood of $H$ given $A$ for a fixed value of $A$
Posterior $=\frac{\text { Prior } \times \text { Likelihood }}{\text { Normalizing Constant }}$
Proof. Conditional probability of $A$ given $H: P(A \mid H)=\frac{P(A \bigcap H)}{P(H)}$
Conditional probability of $H$ given $A: P(H \mid A)=\frac{P(A \bigcap H)}{P(A)}$
Probability of Correctness of Hypothesis $H$ from Observation $A$ $\propto$ Likelihood of A

- Physical Experiments

Likelihood function: $\quad L\langle\theta \mid x\rangle=P\langle x \mid \theta\rangle$
$\backsim$ Probability of getting Observation x from $\theta$ where $x$ : Independent variable \& $\theta$ : Parameter to be determined

Example. $N$ sets of the experiments of the quantity $x$ having Gaussian-type distribution

$$
L\langle\theta \mid x\rangle=\prod_{i=1}^{N} P\left(x_{i} \mid \theta\right) \quad \text { where } x_{i}: \text { Observed value from } i \text {-th experiment }
$$

※ Condition of the maximum likelihood

$$
\left.\frac{\partial L}{\partial \theta}\right|_{\theta=\hat{\theta}}=0 \quad \text { where } \theta=\hat{\theta}: \text { Maximum likelihood estimator }
$$

Goal: Estimate $\mu^{\prime} \& \sigma^{\prime}$ with maximum likelihood
i. From the estimated $\mu^{\prime}, \quad P\left(x_{i}, \mu^{\prime}\right)=\frac{1}{\sigma_{i} \sqrt{2 \pi}} \exp \left(-\frac{\left(x_{i}-\mu^{\prime}\right)^{2}}{2 \sigma_{i}^{2}}\right)$

$$
\text { - } \sigma_{i}: \text { Standard deviation at } i \text {-th experiment }
$$

ii. Check the likelihood of $\mu^{\prime}$

$$
L\left\langle\mu^{\prime} \mid x\right\rangle=\prod_{i=1}^{N} P\left(x_{i}, \mu\right)=\left(\frac{1}{\sigma_{i} \sqrt{2 \pi}}\right)^{N} \exp \left[-\frac{1}{2} \sum_{i=1}^{N}\left(\frac{\left(x_{i}-\mu^{\prime}\right)}{\sigma_{i}}\right)^{2}\right]
$$

: Probability of getting all of the $x_{i}$ from the hypothesized $\mu^{\prime}$

- Maximum Likelihood function

Minimize the exponent $X=\frac{1}{2} \sum_{i=1}^{N}\left(\frac{\left(x_{i}-\mu^{\prime}\right)}{\sigma_{i}}\right)^{2}$
$\frac{\partial X}{\partial \mu^{\prime}}=0 \quad$ (Condition of Maximum likelihood)
(ex. $\mu^{\prime}=\frac{\sum_{i=1}^{N} x_{i}}{N}=\bar{x}$ in case all $\sigma_{i}$ are equal)

- Error on $\mu^{\prime}$

$$
\sigma_{\mu^{\prime}}=\sum_{i=1}^{N}\left[\sigma_{i}^{2}\left(\frac{\partial \mu^{\prime}}{\partial x_{i}}\right)^{2}\right] \quad\left(\sigma_{\mu^{\prime}}=\frac{\sigma^{2}}{N} \text { in case all } \sigma_{i} \text { are equal }\right)
$$

O Accuracy of Theoretical Hypotheses from the sample experiment

- The $\chi^{2}$ Test
- More than one measurement to know if the data follow the theory or a formulated hypothesis (Data fitting, etc.)
- Example: Measurement of length $l$ of a metal bar as a function of temperature $T$.

Step 1. Obtained Data

| $T_{i}\left({ }^{\circ} \mathrm{C}\right)$ | 15 | 17 | 34 | 43 | 58 | 61 | 64 | 70 | 98 |
| :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- | :--- |
| $\bar{l}_{i}(\mathrm{~cm})$ | 1.6 | 2.5 | 3.6 | 4.8 | 5.2 | 6.6 | 7.2 | 8.4 | 9.8 |

Step 2. Formulated Hypothesis:
The data follows a linear function, $\quad y=a x+b \quad(l=a T+b)$


Step 3. Find the best value of explicit independent $a$ and $b$

- Likelihood function

Probability of getting $l_{i}$ from hypothesized $f(T)$

$$
\begin{aligned}
& P_{T_{i}}\left(l_{i}\right)=\left(\frac{1}{\sigma_{i} \sqrt{2 \pi}}\right) \exp \left[-\frac{1}{2}\left(\frac{l_{i}-f\left(T_{i}\right)}{\sigma_{i}}\right)^{2}\right] \\
& \text { where } f\left(T_{i}\right)=a T_{i}+b: \quad \begin{array}{l}
\text { Hypothesis } \\
l_{i}:
\end{array} \quad \begin{array}{ll}
\text { Observed value }
\end{array}
\end{aligned}
$$

Explicitly independent variables in likelihood function: $a$ and $b$,

$$
L\langle a, b \mid l\rangle=P\langle l \mid a, b\rangle=\left(\frac{1}{\sigma_{i} \sqrt{2 \pi}}\right)^{N} \exp \left[-\frac{1}{2} \sum_{i=1}^{N}\left(\frac{l_{i}-f\left(T_{i}\right)}{\sigma_{i}}\right)^{2}\right]
$$

※ Condition of Maximum Likelihood

$$
\sum_{i=1}^{N}\left(\frac{l_{i}-f\left(T_{i}\right)}{\sigma_{i}}\right)^{2}: \text { Minimum }
$$

We define

$$
\chi^{2}=\sum_{i=1}^{N}\left(\frac{y_{i}-f\left(x_{i}\right)}{\sigma_{i}}\right)^{2}=\sum_{i=1}^{N} \chi_{i}^{2} \quad \text { where } \chi_{i}^{2}=\left[\frac{y_{i}-f\left(x_{i}\right)}{\sigma_{i}}\right]^{2}
$$

: Minimize $\chi^{2}$ to maximize the probability

## - The least square deviation

- Goodness-of-fit indicator

○ Determination of $a$ and $b$ by minimizing $\chi^{2}$

$$
\begin{aligned}
& \frac{\partial \chi^{2}}{\partial a}=\frac{\partial}{\partial a}\left[\sum_{i=1}^{N}\left(\frac{y_{i}-f\left(x_{i}\right)}{\sigma_{i}}\right)^{2}\right]=0 \\
& \frac{\partial \chi^{2}}{\partial b}=\frac{\partial}{\partial b}\left[\sum_{i=1}^{N}\left(\frac{y_{i}-f\left(x_{i}\right)}{\sigma_{i}}\right)^{2}\right]=0 \\
& \therefore \quad a=\frac{1}{\delta}\left(\sum_{i=1}^{N} \frac{x_{i}^{2}}{\sigma_{i}^{2}} \sum_{i=1}^{N} \frac{y_{i}}{\sigma_{i}^{2}}-\sum_{i=1}^{N} \frac{x_{i}}{\sigma_{i}^{2}} \sum_{i=1}^{N} \frac{x_{i} y_{i}}{\sigma_{i}^{2}}\right) \\
& b=\frac{1}{\delta}\left(\sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} \sum_{i=1}^{N} \frac{x_{i} y_{i}}{\sigma_{i}^{2}}-\sum_{i=1}^{N} \frac{x_{i}}{\sigma_{i}^{2}} \sum_{i=1}^{N} \frac{y_{i}}{\sigma_{i}^{2}}\right) \\
& \quad \text { where } \delta=\sum_{i=1}^{N} \frac{1}{\sigma_{i}^{2}} \sum_{i=1}^{N} \frac{x_{i}^{2}}{\sigma_{i}^{2}}-\left(\sum_{i=1}^{N} \frac{x_{i}}{\sigma_{i}}\right)^{2}
\end{aligned}
$$

and the uncertainties $\sigma_{a}$ and $\sigma_{b}$

$$
\sigma_{a}^{2}=\sum_{i=1}^{N}\left[\sigma_{i}^{2}\left(\frac{\partial a}{\partial y_{i}}\right)^{2}\right] \& \quad \sigma_{b}^{2}=\left[\sum_{i=1}^{N} \sigma_{i}^{2}\left(\frac{\partial b}{\partial y_{i}}\right)^{2}\right]
$$

※ Mean value of $\chi^{2}=N-m=v \quad$ where $m$ : Number of fitted parameters © $\quad v=N-m$ : Number of degrees of freedom

We define the reduced $\chi_{r}{ }^{2}=\frac{\chi^{2}}{v}$
(8) Mean value of $\chi_{r}{ }^{2}$ : Always 1

