Suppose we are trying to measure the true value of some quantity \((x_T)\).

- We make repeated measurements of this quantity \(\{x_1, x_2, \ldots, x_n\}\).
- The standard way to estimate \(x_T\) from our measurements is to calculate the mean value:
  \[ \bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i \]
- set \(x_T = \bar{x}\).

Does this procedure make sense???

**MLM:** a general method for estimating parameters of interest from data.

**Statement of the Maximum Likelihood Method**

- Assume we have made \(N\) measurements of \(x\) \(\{x_1, x_2, \ldots, x_n\}\).
- Assume we know the probability distribution function that describes \(x\): \(f(x, \square)\).
- Assume we want to determine the parameter \(\square\).
  - **MLM:** pick \(\square\) to maximize the probability of getting the measurements (the \(x_i\)'s) that we did!

How do we use the MLM?

- The probability of measuring \(x_1\) is \(f(x_1, \square)dx\)
- The probability of measuring \(x_2\) is \(f(x_2, \square)dx\)
- The probability of measuring \(x_n\) is \(f(x_n, \square)dx\)
- If the measurements are independent, the probability of getting the measurements we did is:
  \[ L = f(x_1, \square)dx \cdot f(x_2, \square)dx \cdots f(x_n, \square)dx = f(x_1, \square) \cdot f(x_2, \square) \cdots f(x_n, \square)dx^n \]
- We can drop the \(dx^n\) term as it is only a proportionality constant

**Likelihood Function**

\[ L = \prod_{i=1}^{N} f(x_i, \square) \]
We want to pick the $a$ that maximizes $L$:
\[
\frac{\partial L}{\partial a} \bigg|_{a = a^*} = 0
\]
Both $L$ and $\ln L$ have maximum at the same location.
- maximize $\ln L$ rather than $L$ itself because $\ln L$ converts the product into a summation.
- new maximization condition:
\[
\frac{\partial \ln L}{\partial a} \bigg|_{a = a^*} = \frac{1}{N} \sum_{i=1}^{N} \frac{\partial}{\partial a} \ln f(x_i, a) = 0
\]
- $a$ could be an array of parameters (e.g. slope and intercept) or just a single variable.
- equations to determine $a$ range from simple linear equations to coupled non-linear equations.

Example:
- Let $f(x, a)$ be given by a Gaussian distribution.
- Let $a = \bar{a}$ be the mean of the Gaussian.
- We want the best estimate of $a$ from our set of $n$ measurements $\{x_1, x_2, \ldots x_n\}$.
- Let’s assume that $a$ is the same for each measurement.
\[
f(x_i, a) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x_i - \bar{a})^2}{2\sigma^2}}
\]
The likelihood function for this problem is:
\[
L = \prod_{i=1}^{n} f(x_i, a) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{(x_i - \bar{a})^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi \sigma^2}} \prod_{i=1}^{n} e^{-\frac{(x_i - \bar{a})^2}{2\sigma^2}}
\]
Find $\theta$ that maximizes the log likelihood function:

$$\frac{\partial \ln L}{\partial \theta} = \frac{\partial}{\partial \theta} \left[ n \ln \frac{1}{\sqrt{2\pi \sigma}} \sum_{i=1}^{n} (x_i - \theta)^2 \right] = 0$$

$$\frac{\partial}{\partial \theta} \sum_{i=1}^{n} (x_i - \theta)^2 = 0$$

$$\sum_{i=1}^{n} 2(x_i - \theta) = 0$$

$$\sum_{i=1}^{n} x_i = n \theta$$

$$\theta = \frac{1}{n} \sum_{i=1}^{n} x_i$$

Average

If $\sigma_i$ are different for each data point
- $\theta$ is just the weighted average:

$$\theta = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} \sigma_i^2}$$

Weighted average
Example

- Let \( f(x, \theta) \) be given by a Poisson distribution.
- Let \( \theta = \mu \) be the mean of the Poisson.
- We want the best estimate of \( \theta \) from our set of \( n \) measurements \( \{x_1, x_2, \ldots x_n\} \).
- The likelihood function for this problem is:

\[
L = \prod_{i=1}^{n} f(x_i, \theta) = \prod_{i=1}^{n} e^{\theta x_i} \frac{\theta^{x_i}}{x_i!} = \prod_{i=1}^{n} e^{\theta x_i} \frac{\theta^{x_i}}{x_i!} \cdot e^{\theta x_j} \frac{\theta^{x_j}}{x_j!} \cdot \ldots \cdot e^{\theta x_p} \frac{\theta^{x_p}}{x_p!} = e^{n \theta \sum_{i=1}^{n} x_i}
\]

- Find \( \theta \) that maximizes the log likelihood function:

\[
\frac{d \ln L}{d \theta} = \frac{d}{d \theta} \left( n \theta + \ln \left( \prod_{i=1}^{n} x_i \ln(x_1! x_2! \ldots x_n!) \right) \right) = n + \frac{1}{n} \sum_{i=1}^{n} x_i = 0
\]

\[
\theta = \frac{1}{n} \sum_{i=1}^{n} x_i = \text{Average}
\]

Some general properties of the Maximum Likelihood Method

- For large data samples (large \( n \)) the likelihood function, \( L \), approaches a Gaussian distribution.
- Maximum likelihood estimates are usually consistent.
  - For large \( n \) the estimates converge to the true value of the parameters we wish to determine.
- Maximum likelihood estimates are usually unbiased.
  - For all sample sizes the parameter of interest is calculated correctly.
- Maximum likelihood estimate is efficient: the estimate has the smallest variance.
- Maximum likelihood estimate is sufficient: it uses all the information in the observations (the \( x_i \)’s).
- The solution from MLM is unique.
- Bad news: we must know the correct probability distribution for the problem at hand!
Maximum Likelihood Fit of Data to a Function

Suppose we have a set of $n$ measurements:

\[ x_1, y_1 \pm \sigma_1 \]
\[ x_2, y_2 \pm \sigma_2 \]
\[ \ldots \]
\[ x_n, y_n \pm \sigma_n \]

Assume each measurement error ($\sigma$) is a standard deviation from a Gaussian pdf.

Assume that for each measured value $y$, there’s an $x$ which is known exactly.

Suppose we know the functional relationship between the $y$’s and the $x$’s:

\[ y = q(x, \sigma, \sigma, \ldots) \]

$\sigma$, $\sigma$...are parameters.

MLM gives us a method to determine $\sigma$, $\sigma$... from our data.

Example: Fitting data points to a straight line:

\[ q(x, \sigma, \sigma, \ldots) = \sigma + \sigma x \]

\[ L = \prod_{i=1}^{n} f(x_i, \sigma, \sigma) = \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - q(x_i, \sigma, \sigma))^2}{2\sigma_i^2}} = \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - a - b x_i)^2}{2\sigma_i^2}} \]

Find $\sigma$ and $\sigma$ by maximizing the likelihood function $L$ likelihood function:

\[ \frac{\partial \ln L}{\partial \sigma} = \frac{\partial}{\partial \sigma} \left[ \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - q(x_i, \sigma, \sigma))^2}{2\sigma_i^2}} \right] = \frac{\partial}{\partial \sigma} \left[ \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - a - b x_i)^2}{2\sigma_i^2}} \right] = 0 \]

\[ \frac{\partial \ln L}{\partial \sigma} = \frac{\partial}{\partial \sigma} \left[ \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - q(x_i, \sigma, \sigma))^2}{2\sigma_i^2}} \right] = \frac{\partial}{\partial \sigma} \left[ \prod_{i=1}^{n} \frac{1}{\sigma_i \sqrt{2\pi}} e^{-\frac{(y_i - a - b x_i)^2}{2\sigma_i^2}} \right] = 0 \]

two linear equations with two unknowns
Assume all \( s \)’s are the same for simplicity:
\[
\sum_{i=1}^{n} y_i x_i = 0 \\
\sum_{i=1}^{n} y_i x_i^2 = 0
\]
We now have two equations that are linear in the two unknowns, \( a \) and \( b \).
\[
\begin{align*}
\sum_{i=1}^{n} y_i &= n a + \sum_{i=1}^{n} x_i \\
\sum_{i=1}^{n} y_i x_i &= \sum_{i=1}^{n} x_i + \sum_{i=1}^{n} x_i^2
\end{align*}
\]

We will see this problem again when we talk about “least squares” (“chi-square”) fitting.

EXAMPLE:
A trolley moves along a track at constant speed. Suppose the following measurements of the time vs. distance were made. From the data find the best value for the speed \( v \) of the trolley.

<table>
<thead>
<tr>
<th>Time ( t ) (seconds)</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
<th>5.0</th>
<th>6.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Distance ( d ) (mm)</td>
<td>11</td>
<td>19</td>
<td>33</td>
<td>40</td>
<td>49</td>
<td>61</td>
</tr>
</tbody>
</table>

Our model of the motion of the trolley tells us that:
\[
d = d_0 + vt
\]
We want to find \( v \), the slope of the straight line describing the motion of the trolley. We need to evaluate the sums listed in the above formula:

\[
\begin{align*}
\sum_{i=1}^{n} x_i &= \sum_{i=1}^{6} t_i = 21 \text{ s} \\
\sum_{i=1}^{n} y_i &= \sum_{i=1}^{6} d_i = 213 \text{ mm} \\
\sum_{i=1}^{n} x_i y_i &= \sum_{i=1}^{6} t_i d_i = 919 \text{ s} \cdot \text{mm} \\
\sum_{i=1}^{n} x_i^2 &= \sum_{i=1}^{6} t_i^2 = 91 \text{ s}^2 \\
\end{align*}
\]

\[
v = \frac{n \sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n \sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2} = \frac{6 \times 919 - 21 \times 213}{6 \times 91 - 21^2} = 9.9 \text{ mm/s}
\]

best estimate of the speed

\[d_0 = 0.8 \text{ mm}\]

best estimate of the starting point
MLM fit to the data for $d = d_0 + vt$

- The line **best** represents our data.
- Not all the data points are "on" the line.
- The line minimizes the sum of squares of the deviations between the line and our data ($d_i$):
  $$
  \Box = \sum_{i=1}^{n} \left(d_i - \text{prediction}_i\right)^2 = \sum_{i=1}^{n} \left(d_i - (d_0 + vt_i)\right)^2
  $$

K.K. Gan

L5: Maximum Likelihood Method