Lecture 5
Maximum Likelihood Method

- 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:
    \[
    \mu_x = \frac{1}{N} \sum_{i=1}^{N} x_i
    \]
  - set \(x_T = \mu_x\).
  - **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 \(x\): \(f(x, \alpha)\).
  - Assume we want to determine the parameter \(\alpha\).
    - **MLM**: pick \(\alpha\) 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 \(f(x_1, \alpha)dx\)
  - The probability of measuring \(x_2\) is \(f(x_2, \alpha)dx\)
  - The probability of measuring \(x_n\) is \(f(x_n, \alpha)dx\)
  - If the measurements are independent, the probability of getting the measurements we did is:
    \[
    L = f(x_1, \alpha)dx \cdot f(x_2, \alpha)dx \cdots f(x_n, \alpha)dx = f(x_1, \alpha) \cdot f(x_2, \alpha) \cdots f(x_n, \alpha)dx^n
    \]
  - We can drop the \(dx^n\) term as it is only a proportionality constant
  - **Likelihood Function**
    \[
    L = \prod_{i=1}^{N} f(x_i, \alpha)
    \]
We want to pick the $\alpha$ that maximizes $L$:

$$\frac{\partial L}{\partial \alpha_{\alpha=\alpha^*}} = 0$$

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:

$$\frac{\partial \ln L}{\partial \alpha_{\alpha=\alpha^*}} = \sum_{i=1}^{N} \frac{\partial \ln f(x_i, \alpha)}{\partial \alpha_{\alpha=\alpha^*}} = 0$$

- $\alpha$ could be an array of parameters (e.g. slope and intercept) or just a single variable.
- equations to determine $\alpha$ range from simple linear equations to coupled non-linear equations.

- Example:

  - Let $f(x, \alpha)$ be given by a Gaussian distribution.
  - Let $\alpha = \mu$ be the mean of the Gaussian.
  - We want the best estimate of $\alpha$ from our set of $n$ measurements $\{x_1, x_2, \ldots x_n\}$.
  - Let’s assume that $\sigma$ is the same for each measurement.

$$f(x_i, \alpha) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i-\alpha)^2}{2\sigma^2}}$$

- The likelihood function for this problem is:

$$L = \prod_{i=1}^{n} f(x_i, \alpha) = \prod_{i=1}^{n} \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_i-\alpha)^2}{2\sigma^2}} = \left[ \frac{1}{\sigma \sqrt{2\pi}} \right]^{n} e^{-\frac{(x_1-\alpha)^2}{2\sigma^2}} e^{-\frac{(x_2-\alpha)^2}{2\sigma^2}} \ldots e^{-\frac{(x_n-\alpha)^2}{2\sigma^2}} = \left[ \frac{1}{\sigma \sqrt{2\pi}} \right]^{n} e^{-\sum_{i=1}^{n} \frac{(x_i-\alpha)^2}{2\sigma^2}}$$
Find $\alpha$ that maximizes the log likelihood function:

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

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

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

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

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} x_i$$  \text{Average}

If $\sigma$ are different for each data point

$\alpha$ is just the weighted average:

$$\alpha = \frac{\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$$  \text{Weighted average}
Example

Let $f(x, \alpha)$ be given by a Poisson distribution.

Let $\alpha = \mu$ be the mean of the Poisson.

We want the best estimate of $\alpha$ from our set of $n$ measurements $\{x_1, x_2, \ldots, x_n\}$.

The likelihood function for this problem is:

$$L = \prod_{i=1}^{n} f(x_i, \alpha) = \prod_{i=1}^{n} \frac{e^{-\alpha} \alpha^{x_i}}{x_i!} = \frac{e^{-\alpha} \alpha^{x_1}}{x_1!} \frac{e^{-\alpha} \alpha^{x_2}}{x_2!} \cdots \frac{e^{-\alpha} \alpha^{x_n}}{x_n!} = e^{-n\alpha} \frac{\alpha^{\sum x_i}}{x_1!x_2!\cdots x_n!}$$

Find $\alpha$ that maximizes the log likelihood function:

$$\frac{d \ln L}{d \alpha} = \frac{d}{d \alpha} \left( -n\alpha + \ln \alpha \cdot \sum_{i=1}^{n} x_i - \ln(x_1!x_2!\cdots x_n!) \right) = -n + \frac{1}{\alpha} \sum_{i=1}^{n} x_i = 0$$

$$\alpha = \frac{1}{n} \sum_{i=1}^{n} x_i \quad \text{Average}$$

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:
  \[
  y = q(x, \alpha, \beta, \ldots)
  \]
  \( \alpha, \beta \ldots \) are parameters.
- MLM gives us a method to determine \( \alpha, \beta \ldots \) from our data.
- Example: Fitting data points to a straight line:
  \[
  q(x, \alpha, \beta, \ldots) = \alpha + \beta x
  \]

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

- Find \( \alpha \) and \( \beta \) by maximizing the likelihood function \( L \) likelihood function:

\[
\frac{\partial \ln L}{\partial \alpha} = \frac{\partial}{\partial \alpha} \sum_{i=1}^{n} \ln \left( \frac{1}{\sigma_i \sqrt{2\pi}} \right) - \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_i^2} = \sum_{i=1}^{n} -\frac{2(y_i - \alpha - \beta x_i)(-1)}{2\sigma_i^2} = 0
\]

\[
\frac{\partial \ln L}{\partial \beta} = \frac{\partial}{\partial \beta} \sum_{i=1}^{n} \ln \left( \frac{1}{\sigma_i \sqrt{2\pi}} \right) - \frac{(y_i - \alpha - \beta x_i)^2}{2\sigma_i^2} = \sum_{i=1}^{n} -\frac{2(y_i - \alpha - \beta x_i)(-x_i)}{2\sigma_i^2} = 0
\]

\( \left[ \begin{array}{c}
\text{two linear equations with two unknowns}
\end{array} \right] \)
Assume all $\sigma$'s are the same for simplicity:

$$\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} \alpha - \sum_{i=1}^{n} \beta x_i = 0$$
$$\sum_{i=1}^{n} y_i x_i - \sum_{i=1}^{n} \alpha x_i - \sum_{i=1}^{n} \beta x_i^2 = 0$$

We now have two equations that are linear in the two unknowns, $\alpha$ and $\beta$.

$$\sum_{i=1}^{n} y_i = n\alpha + \beta \sum_{i=1}^{n} x_i$$
$$\sum_{i=1}^{n} y_i x_i = \alpha \sum_{i=1}^{n} x_i + \beta \sum_{i=1}^{n} x_i^2$$

Matrix form

$$\begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} y_i x_i \end{bmatrix} = \begin{bmatrix} n & \sum_{i=1}^{n} x_i \\ \sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$$

Taylor Eqs. 8.10-12

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 velocity ($v$) 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 ($\beta$) of the straight line describing the motion of the trolley. We need to evaluate the sums listed in the above formula:

\[
\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} t_i = 21 \text{ s}
\]

\[
\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} d_i = 213 \text{ mm}
\]

\[
\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} t_i d_i = 919 \text{ s} \cdot \text{mm}
\]

\[
\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} t_i^2 = 91 \text{ s}^2
\]

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

(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 ($d_i$):
  \[ \delta = \sum_{i=1}^{n} (\text{data}_i - \text{prediction}_i)^2 = \sum_{i=1}^{n} (d_i - (d_0 + vt_i))^2 \]  

Least square fit