Lecture 6
Chi Square Distribution ($c^2$) and Least Squares Fitting

Chi Square Distribution ($c^2$)

Suppose:
- We have a set of measurements $\{x_1, x_2, \ldots, x_n\}$.
- We know the true value of each $x_i$ ($x_{t1}, x_{t2}, \ldots, x_{tn}$).
    - We would like some way to measure how good these measurements really are.
- Obviously the closer the ($x_1, x_2, \ldots, x_n$)'s are to the ($x_{t1}, x_{t2}, \ldots, x_{tn}$)'s,
  - the better (or more accurate) the measurements.
  - can we get more specific?

Assume:
- The measurements are independent of each other.
- The measurements come from a Gaussian distribution.
- ($\sigma_1, \sigma_2, \ldots, \sigma_n$) be the standard deviation associated with each measurement.

Consider the following two possible measures of the quality of the data:

\[
R \equiv \prod_{i=1}^{n} \frac{x_i - x_{ti}}{\sigma_i}
\]

\[
c^2 \equiv \prod_{i=1}^{n} \frac{(x_i - x_{ti})^2}{\sigma_i^2}
\]

Which of the above gives more information on the quality of the data?
- Both $R$ and $c^2$ are zero if the measurements agree with the true value.
- $R$ looks good because via the Central Limit Theorem as $n \rightarrow$ the sum $\rightarrow$ Gaussian.
- However, $c^2$ is better!
One can show that the probability distribution for $\chi^2$ is exactly:

$$p(\chi^2, n) = \frac{1}{2^{n/2} \Gamma(n/2)} [\chi^2]^{n/2} e^{-\chi^2/2} \quad 0 \leq \chi^2 \leq \infty$$

This is called the "Chi Square" ($\chi^2$) distribution.

$\Gamma(x)$ is the Gamma Function:

$$\Gamma(x) = \int_0^\infty e^{-t} t^{x-1} dt \quad x > 0$$

$$\Gamma(n+1) = n! \quad n = 1, 2, 3...$$

$$\Gamma(\frac{1}{2}) = \sqrt{\pi}$$

This is a continuous probability distribution that is a function of two variables:

$\chi^2$

Number of degrees of freedom (dof):

$n = \# \text{ of data points} - \# \text{ of parameters calculated from the data points}$

Example: We collected N events in an experiment.

- We histogram the data in $n$ bins before performing a fit to the data points.
- We have $n$ data points!

Example: We count cosmic ray events in 15 second intervals and sort the data into 5 bins:

<table>
<thead>
<tr>
<th>Number of counts in 15 second intervals</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of intervals</td>
<td>2</td>
<td>7</td>
<td>6</td>
<td>3</td>
<td>2</td>
</tr>
</tbody>
</table>

we have a total of 36 cosmic rays in 20 intervals
we have only 5 data points
Suppose we want to compare our data with the expectations of a Poisson distribution:

$$N = N_0 \frac{e^{-\lambda} \lambda^m}{m!}$$
Since we set $N_0 = 20$ in order to make the comparison, we lost one degree of freedom:

$$n = 5 - 1 = 4$$

If we calculate the mean of the Poission from data, we lost another degree of freedom:

$$n = 5 - 2 = 3$$

Example: We have 10 data points.

- Let $\mu$ and $\sigma$ be the mean and standard deviation of the data.
- If we calculate $\mu$ and $\sigma$ from the 10 data point then $n = 8$.
- If we know $\mu$ and calculate $\sigma$ then $n = 9$.
- If we know $\mu$ and calculate $\sigma$ then $n = 9$.
- If we know $\mu$ and calculate $\sigma$ then $n = 10$.

Like the Gaussian probability distribution, the probability integral cannot be done in closed form:

$$P(\chi^2 > a) = \int_a^{\infty} p(\chi^2, n) d\chi^2 = \frac{1}{\sqrt{2^nn!}} \int_a^{\infty} e^{-\chi^2/2} \chi^{n/2-1} d\chi^2$$

We must use to a table to find out the probability of exceeding certain $\chi^2$ for a given dof.

For $n \geq 20$, $P(\chi^2 > a)$ can be approximated using a Gaussian pdf with $a = (2\chi^2)^{1/2} - (2n-1)^{1/2}$
Example: What’s the probability to have $\chi^2 > 10$ with the number of degrees of freedom $n = 4$?
- Using Table D of Taylor we find $P(\chi^2 > 10, n = 4) = 0.04$.
- We say that the probability of getting a $\chi^2 > 10$ with 4 degrees of freedom by chance is 4%.

Some not so nice things about the $\chi^2$ distribution:
- Given a set of data points two different functions can have the same value of $\chi^2$.
  - Does not produce a unique form of solution or function.
- Does not look at the order of the data points.
  - Ignores trends in the data points.
- Ignores the sign of differences between the data points and “true” values.
  - Use only the square of the differences.
- There are other distributions/statistical test that do use the order of the points:
  “run tests” and “Kolmogorov test”
Least Squares Fitting

Suppose we have \( n \) data points \((x_i, y_i, s_i)\).

Assume that we know a functional relationship between the points,
\[ y = f(x, a, b, \ldots) \]
Assume that for each \( y_i \) we know \( x_i \) exactly.
The parameters \( a, b, \ldots \) are constants that we wish to determine from our data points.
A procedure to obtain \( a \) and \( b \) is to minimize the following \( \mathcal{F} \) with respect to \( a \) and \( b \).
\[
\mathcal{F}^2 = \sum_{i=1}^{n} \left[ y_i - f(x_i, a, b) \right]^2 \sum_{i=1}^{n} s_i^2
\]
This is very similar to the Maximum Likelihood Method.
For the Gaussian case MLM and LS are identical.
Technically this is a \( \mathcal{F} \) distribution only if the \( y \)'s are from a Gaussian distribution.
Since most of the time the \( y \)'s are not from a Gaussian we call it “least squares” rather than \( \mathcal{F} \).

Example: We have a function with one unknown parameter:
\[ f(x, b) = 1 + bx \]
Find \( b \) using the least squares technique.

We need to minimize the following:
\[
\mathcal{F}^2 = \sum_{i=1}^{n} \left[ y_i - f(x_i, a, b) \right]^2 \sum_{i=1}^{n} s_i^2 = \sum_{i=1}^{n} \left[ y_i - 1 - bx_i \right]^2 \sum_{i=1}^{n} s_i^2
\]
To find the \( b \) that minimizes the above function, we do the following:
\[
\frac{\partial \mathcal{F}^2}{\partial b} = \sum_{i=1}^{n} \frac{\partial}{\partial b} \left[ y_i - 1 - bx_i \right] x_i = 0
\]
\[
\sum_{i=1}^{n} y_i x_i \sum_{i=1}^{n} s_i^2 = \sum_{i=1}^{n} x_i \sum_{i=1}^{n} s_i^2 = \sum_{i=1}^{n} bx_i^2 = 0
\]
Each measured data point \( (y_i) \) is allowed to have a different standard deviation \( (\sigma_i) \).

LS technique can be generalized to two or more parameters for simple and complicated (e.g. non-linear) functions.

One especially nice case is a polynomial function that is linear in the unknowns \( (a_i) \):

\[
f(x,a_1...a_n) = a_1 + a_2 x + a_3 x^2 + a_n x^n
\]

We can always recast problem in terms of solving \( n \) simultaneous linear equations.

We use the techniques from linear algebra and invert an \( n \times n \) matrix to find the \( a_i \)'s!

Example: Given the following data perform a least squares fit to find the value of \( b \).

\[
f(x,b) = 1 + bx
\]

<table>
<thead>
<tr>
<th>( x )</th>
<th>1.0</th>
<th>2.0</th>
<th>3.0</th>
<th>4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>( y )</td>
<td>2.2</td>
<td>2.9</td>
<td>4.3</td>
<td>5.2</td>
</tr>
<tr>
<td>( \sigma )</td>
<td>0.2</td>
<td>0.4</td>
<td>0.3</td>
<td>0.1</td>
</tr>
</tbody>
</table>

Using the above expression for \( b \) we calculate:

\[
b = 1.05
\]
A plot of the data points and the line from the least squares fit:

If we assume that the data points are from a Gaussian distribution, we can calculate a $\chi^2$ and the probability associated with the fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{(y_i - (1.05x_i))^2}{\sigma_i^2} = \frac{2.2}{0.2} + \frac{2.9}{0.4} + \frac{4.3}{0.3} + \frac{5.2}{0.1} = 1.04$$

From Table D of Taylor:
- The probability to get $\chi^2 > 1.04$ for 3 degrees of freedom $\approx 80\%$.
- We call this a "good" fit since the probability is close to 100%.
- If however the $\chi^2$ was large (e.g. 15),
  - the probability would be small ($\approx 0.2\%$ for 3 dof).
  - we say this was a "bad" fit.

RULE OF THUMB:
A "good" fit has $\chi^2 / \text{dof} \leq 1$