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

Chi Square Distribution (χ^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:

$$R \equiv \sum_{i=1}^{n} \frac{x_i - x_{ti}}{\sigma_i}$$

$$\chi^2 \equiv \sum_{i=1}^{n} \frac{(x_i - x_{ti})^2}{\sigma_i^2}$$

  - Which of the above gives more information on the quality of the data?
    - Both R and χ^2 are zero if the measurements agree with the true value.
    - R looks good because via the Central Limit Theorem as $n \to \infty$ the sum $\to$ Gaussian.
    - However, χ^2 is better!

- One can show that the probability distribution for χ^2 is exactly:

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

  - This is called the "Chi Square" (χ^2) distribution.
    - Γ is the Gamma Function:

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

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

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

  - This is a continuous probability distribution that is a function of two variables:
    - χ^2
    - Number of degrees of freedom (dof):

      $n =$ # of data points - # 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:

Number of counts in 15 second intervals	0	1	2	3	4
Number of intervals	2	7	6	3	2

        - 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:

$$N=N_0\frac{e^{-\mu}\mu^m}{m!}$$

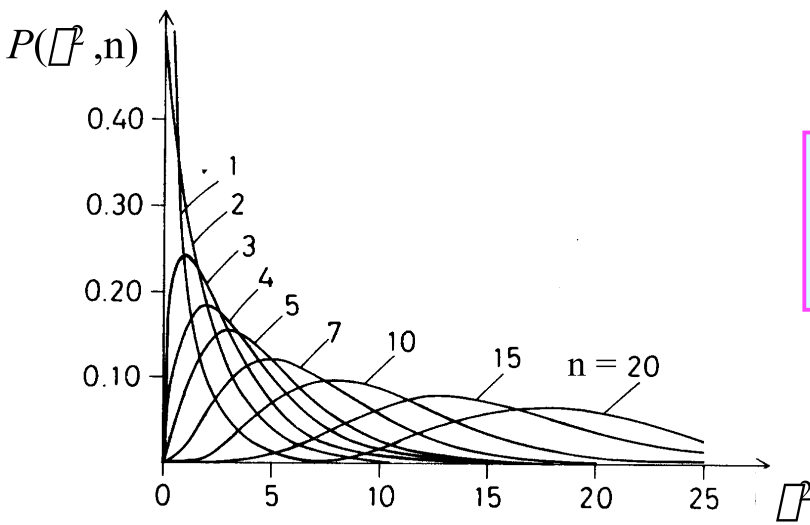
- Since we set $N_0 = 20$ in order to make the comparison, we lost one degree of freedom:
  $n = 5 - 1 = 4$
- If we calculate the mean of the Poission from data, we lost another degree of freedom:
  $n = 5 - 2 = 3$

- Example: We have 10 data points.
  - Let μ and σ be the mean and standard deviation of the data.
  - If we calculate μ and σ from the 10 data point then $n = 8$.
  - If we know μ and calculate σ then $n = 9$.
  - If we know σ and calculate μ then $n = 9$.
  - If we know μ and σ then $n = 10$.

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

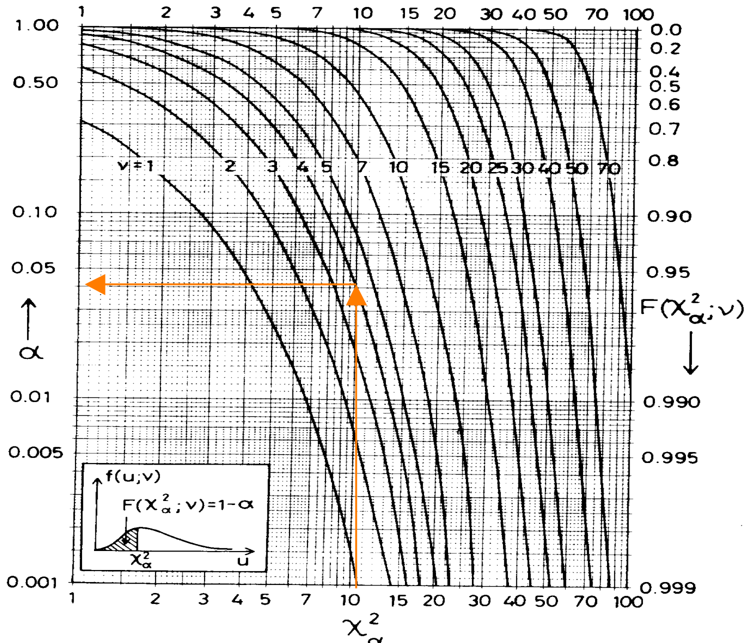
$$P(\chi^2 > a) = \int_a^\infty p(\chi^2, n)d\chi^2 = \int_a^\infty \frac{1}{2^{n/2}\Gamma(n/2)}[\chi^2]^{n/2-1}e^{-\chi^2/2}d\chi^2$$

- We must use to a table to find out the probability of exceeding certain χ^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 χ^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 χ^2 distribution:
  - Given a set of data points two different functions can have the same value of χ^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, σ_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, … are constants that we wish to determine from our data points.
  - A procedure to obtain a and b is to minimize the following χ^2 with respect to a and b.
    $$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - f(x_i, a, b)]^2}{\sigma_i^2}$$
    - This is very similar to the Maximum Likelihood Method.
      - For the Gaussian case MLM and LS are identical.
      - Technically this is a χ^2 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 χ^2.
- 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:
    $$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - f(x_i, a, b)]^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{[y_i - 1 - bx_i]^2}{\sigma_i^2}$$
  - To find the b that minimizes the above function, we do the following:
    $$\frac{\partial \chi^2}{\partial b} = \frac{\partial}{\partial b} \sum_{i=1}^{n} \frac{[y_i - 1 - bx_i]^2}{\sigma_i^2} = \sum_{i=1}^{n} \frac{-2[y_i - 1 - bx_i]x_i}{\sigma_i^2} = 0$$
    $$\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{b x_i^2}{\sigma_i^2} = 0$$

$$b = \frac{\sum_{i=1}^{n} \frac{y_i x_i}{\sigma_i^2} - \sum_{i=1}^{n} \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}}$$

- Each measured data point (y_i) is allowed to have a different standard deviation (σ_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):

    $f(x, a_1 \ldots a_n) = a_1 + a_2 x + a_3 x^2 + a_n x^{n-1}$
    - We can always recast problem in terms of solving n simultaneous linear equations.
      - ☞ We use the techniques from linear algebra and invert an n x n matrix to find the a_i's!

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

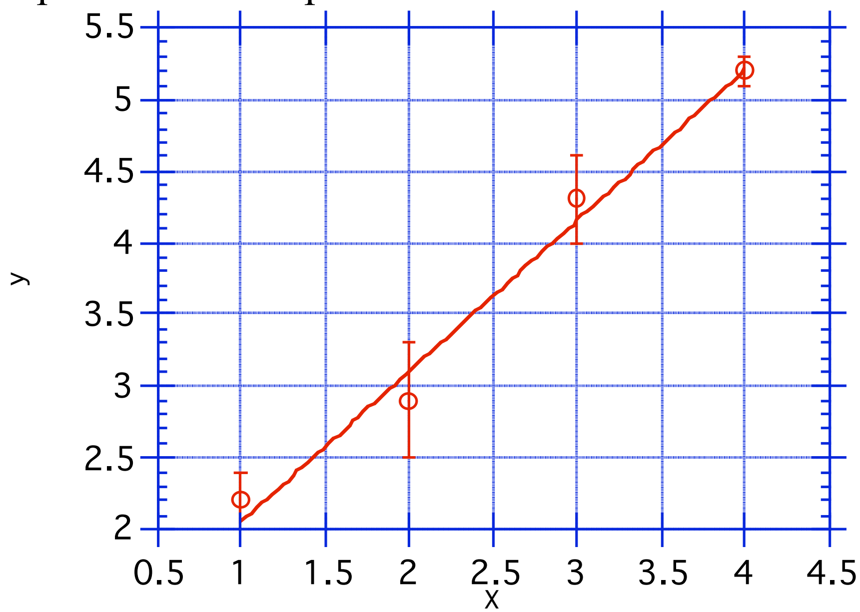
  $f(x, b) = 1 + bx$

x	1.0	2.0	3.0	4.0
y	2.2	2.9	4.3	5.2
σ	0.2	0.4	0.3	0.1

  - 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 <u>assume</u> that the data points are from a Gaussian distribution,
  - we can calculate a χ^2 and the probability associated with the fit.

$$\chi^2 = \sum_{i=1}^{n} \frac{[y_i - 1 - 1.05x_i]^2}{\sigma_i^2} = \left(\frac{2.2 - 2.05}{0.2}\right)^2 + \left(\frac{2.9 - 3.1}{0.4}\right)^2 + \left(\frac{4.3 - 4.16}{0.3}\right)^2 + \left(\frac{5.2 - 5.2}{0.1}\right)^2 = 1.04$$

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

RULE OF THUMB
A "good" fit has χ^2 /dof ≤ 1