Data modeling and χ^2 minimization

As physicists we are aware that every time we perform a measurement it comes with an uncertainty. There is a statistical uncertainty, which we can bring down by repeating measurements many times, and a systematic uncertainty, which has to do with how accurate our instruments are.

In this activity we will focus on statistical uncertainties. The activity is divided in two parts. In the first part we learn how to synthetize a set of measerement in a most probable value and an uncertainty. In the second part we learn how to find the best analytical model for a set of measurements that depend on an independent quantity.

Part I

Consider the set of measurements in the table below. Those are individual measure of the sun's luminosity made by the same instrument at different times.

Measurement number	1	2	3	4	5	6	7	8	9	10
Measured value	9.0	4.3	8.2	9.5	6.4	10.4	8.1	7.3	10.6	6.0

The most propable value of the true Sun's luminosity is:

$$\bar{m} = \frac{\sum m_i}{N} \tag{1}$$

where m_i are the individual measurements and N is the total number of measurements.

The uncertainlty on that measurement is given instead by the standard error:

$$\sigma_m = \frac{\sqrt{\sum (m_i - \bar{m})^2}}{N-1} \tag{2}$$

<u>Activity tasks</u>: write a python code that

1. Creates an array of measurements with the values in the table
2. Computes the most probable value
3. Computes the standard error
4. Prints them out

At this point, call someone from the instructional team as a checkpoint to your work.

Part II

The above measurements were performed at a specific frequency. The instrument frequency can be precisely tuned and a set of measurements taken at various frequencies. After averaging the measurement at each frequency and computing standard errors, the researcher obtains the following measuremnts with uncertainties:

frequency	measurement	uncertainty
1	0.05	0.1
1.5	0.1	0.15
2.	0.0075	0.12
2.5	1.86	0.17
3.	Insert the values	you found in part I
3.5	6.0	0.72
4.	0.41	0.3
4.5	0.03	0.1
5.	0.1	0.13
5.5	0.05	0.11

Activity tasks: continue your code to:

1. create arrays with the frequencies, measurements, and uncertainties
2. make a plot with the measurements as a function of frequency. (hint: you can use plt.errorbar)

The researcher knows that something is going at a frequency about 3. They know spectral features have Gaussian shapes

$$l(\nu) = 10e^{-\frac{(\nu-\nu_0)^2}{2\Delta_\nu^2}} \tag{3}$$

but they cant quite tell what value of central frequency ν_0 nor line width $\Delta\nu$ are implied by the measurements.

The idea here is to find the best Gaussian by minimizing the "distance" between the measurement and the line at the frequencies for which we have made the measurements. The "" are use around the word distance since we use a specific quantity to measure distance that is called χ^2. It is defined as:

$$\chi^2 = \sum \frac{[m_i - l(\nu_i)]^2}{\sigma_i^2} \tag{4}$$

The questions we ask are:

1. Is the Gaussian shape a good description of the data?
2. If so, what are the values of ν_0 and Δ_ν that best describe the data?
3. What are the uncertainties σ_{ν_0} and σ_{Δ_ν} on these values?

To which we answer:

1. Yes, if the minimum possible value of the χ^2 is not bigger than twice the number of data points
2. Those for which the value of the χ^2 is minimum
3. The values of ν_0 and Δ_ν for which $\chi^2(\nu_0, \Delta_\nu) \leq \chi^2_{\text{min}}$

Activity tasks: continue your code to:

1. Map the $\chi^2(\nu_0, \Delta_\nu)$ surface to answer the 3 questions for your data. This can be done by:
   a) Create arrays of guess values for $2.5 < \nu_0 < 3.5$ and $0.25 < \Delta_\nu < 0.4$
   b) Compute the χ^2 value for each pair $(\nu_{0,i}, \Delta_{\nu,j})$ and save those result in an array
   c) Find the pair $(\nu_{0,i}, \Delta_{\nu,j})$ for which the χ^2 is minimum
   d) Ask yourself whether such χ^2 is small enough
   e) Make a contour plot of the χ^2 array that you created in point b. Plot a single contour level for which $\chi^2 = \chi^2_{\text{min}} + 2.3$
2. Add the best model to the plot with the data. Make sure it is visually satisfying

Here is some pseudocode to help with this last part:

- create an array of guess values for the central frequency ν_0.
- create a second array of guess values for the Gaussian width Δ_ν.
- create an empty 2-dimensional array for storing the χ^2 values that you will compute
- use a nested double for loop over the guess ν_0 and Δ_ν
- for each pair of numbers $(\nu_0[i], \Delta_\nu[j])$ compute the Gaussian modelfrom Eq. (3) using the values in the first column of the table above as ν
- still within the double loop, compute the χ^2 from Eq. (4), wiere m_i are the data in the second column of the table and σ_i the uncertainties in the third column
- store the calculated χ^2 value in the 2-dimensional array
- Out of the double loop, search the χ^2 2-dimensional array for its minimum value
- Establish whether it is acceptable ($\chi^2_{\text{min}} \sim \#_{data}$)
- Find the value of ν_0 and Δ_ν for which the minimum χ^2 is attained. These are your best estimates of the central frequency and width of the signal.
- display a contour plot of the χ^2 2-dimensional array. Set the only level to be displayed as $\chi^2_{\text{min}} + 2.3$
- Also plot your two best estimates for ν_0 and Δ_ν as a single point in the same graph. Does the best estimate lay at the center of the ellipse?

Extra Activity

In the **Files** tab on Canvas, look in the ”Meterials” folder. You will find a file ”data.txt”. It contains a set of data organized in a (21×31) elements array. The first column of such array contains the times at which 21 measurements were taken. For each time, the remaining row of data contains 30 measurements of the same quantity. The model that can be used to describe the measurements is:

$$f(t) = \frac{1}{1 + \left(\frac{t}{A}\right)^B} \tag{5}$$

Your goal is similar to the original activity: find the best estimates for the values of A and B and they uncertainty by using the χ^2 method.