Chi-square fitting - 1

Chih-Keng Hung

April 23, 2025

1 Practice 1

Purpose:

1. To test the difference between least square fitting and chi-square fitting using `scipy.optimize.curve_fit`.

Task:

1. Generate 20 sets of data points defined by a function $y = a * x + b + noise$ with `x = np.linspace(1, 10, 10)`.
2. The noise is sampled from a normal distribution with mean $= 0$ and standard deviation $= \sigma_0 = 3$.
3. Change the σ_0 of y at $x = 4, 5, 6, 7$ to $\sigma_0 = 10$.
4. You might see the deviation of the data points at these x values are larger than the others.
5. Get the *mean* and *standard deviation* of the data points.(mean value of y might be close to $a * x + b$)
6. Fit these mean value of y with the function $y = a * x + b$ using `curve_fit`.
7. Compare the differce between *least square fitting method* and *chi-square fitting method* by calculating the χ^2 value.
8. Which fitting method is better in this case?
9. State the reason.

Hint:

1. *"sigma"* parameter in `curve_fit` can be used to assign the standard deviation of the data points.
2. To pass or not to pass the *sigma* parameter might affect the fitting method.
3. Use **"actual"** standard deviation to pass into the *sigma* parameter in `curve_fit` and calculate the chi-square value.
4. After generating the data points, fit the data points with different fitting methods instantly to compare the results or you might use the `np.random.seed` to fix the random number generation process.
5. The data cannot achieve a perfect fit to the model solely due to random errors, as demonstrated in your approach. Not to mention, real-world data is inevitably influenced by systematic errors, further complicating the accuracy of the fit. That's why the ability to estimate the goodness of fit is crucial.

2 Practice 2

Purpose:

1. To understand the basic concept of chi-square distribution using normal distribution.
2. To know the meaning of chi-square value.

Task:

1. Generate an 1-d array with 500 elements, each element is sampled from a normal distribution with mean $= 0$ and standard deviation $= 1$.

2. Calculate the chi-square value of each data point: $chisqr = (data\_point - \mu)^2/\sigma$, you'll find out $\sigma \approx 1, \mu \approx 0$. (Note that `len`(chisqr) = 500 in this case. You might feel confused about the definition of chisqr here, but just accept it for now. You can think about what it probably means in the fitting process you've done before.)

3. Plot the histogram of these chi-square values.

4. Compare the histogram with the chi-square distribution with degree of freedom = 1.

5. Create several (2 to 10) 1-d arrays similar to the ones above, each array contains 500 elements sampled from a normal distribution with mean = 0 and standard deviation = 1. Then calculate the chi-square value of each data point in each array and sum them up to get a new 1-d array that also contains 500 elements representing the sum of chi-square values. (e.g. array1 = [1, 2, 3,...(500^{th} value)], array2 = [4, 5, 6,...(500^{th} value)], sum_array = $(\text{array1}/\sigma 1)^2 + (\text{array2}/\sigma 2)^2$, `len`(sum_array) = 500)

6. Plot the histogram of these chi-square values (sum_array).

7. Compare the histogram with the chi-square distribution with degree of freedom equal to the number of 1-d array(2∼10) you generated.

8. What do you find?

9. Considering your previous work on chi-square fitting, describe the relation between those μ, σ, elements, and the fitting process.

Hint:

1. Eusure the definition of chi-square distribution function is correct.

2. In a loose sense, a normally distributed population can be described by μ and σ since these parameters define its shape and statistical properties.