Chi-square fitting - 3

Chih-Keng Hung

May 12, 2025

1 Practice 1

Purpose:

1. To understand the meaning of p-value.

Task:

1. By using previous knowledge, choose your own function to generate data points.
2. Generate 1000 sets of data points with the same function with normally distributed random noise.
3. Fit these data points with the function you used to generate the data points.
4. Calculate the χ^2 values of each fitting result.
5. Explain how to calculate the **p-value**.
6. Create your own **p-value** calculator to calculate the **p-value** of each χ^2 instead of using functions in `scipy.stats`.
7. Plot the histogram of χ^2.
8. Mark a certain χ^2 value with its corresponding **p-value** on the histogram.
9. Once again, I ask you: Will the fitted parameters be closer to the initial parameters you set to genetate the data points with larger **p-value**? You should trust your results rather than rely on preconceived bias.

Hint:

1. Chi-Square Table: `https://www.medcalc.org/manual/chi-square-table.php`
2. Cumulative distribution function (CDF)

2 Practice 2

Purpose:

1. Be able to state the goodness of fitting result using p-value, χ^2 and significance level.

Task:

1. Using prior knowledge, select an appropriate function with noise to generate your dataset.
2. Plot a histogram of the χ^2 values obtained from fitting each dataset.
3. Mark the **p-value** get from the fitting result of `mean(y)`.
4. You might get a χ^2 of the `mean(y)` fitting result close to zero and a **p-value** close to 1.
5. Also mark 5% significance level, and 95% significance level on the histogram.
6. State the goodness of fitting result by comparing the **p-value** with the significance level. (Notice that you can set any value of significance level you want actually.)
7. If the **p-value** get from the fit is less than 0.05, what does it mean?
8. Make a conclusion how to make a accurate statement about the goodness of fitting result.

Hint:

1. Chi-square test
2. Null hypothesis
3. Do not be messed up with **significance level** and **confidence level**.

Reminder:

Trust the work you have done and the results you have obtained. Question the findings of others, and approach their statements with skepticism. Truth emerges through defense.

All the practice you have done so far falls within the scope of ideal case studies, including the following assumptions:

1. Data points consist of only random error.
2. Error is presented in y-axis only.
3. Error is normally distributed.

Although these assumptions are not always true, they are still useful to help with understanding the concept of fitting. They offer an idealized perspective on the fitting process, facilitating more efficient discussions.