Data Analysis-II

Least sqaure fit and extrapolation

Soon-Hyung Yook

December 4, 2015

Table of Contents

General Description

- Based on some theoretical grounds or assumption
  - if we know that a quantity y is a function of x with m parameters $a_1, a_2, \cdots, a_m$.
  - $\Rightarrow y = f(a_1, a_2, \cdots, a_m; x)$.
- $\Rightarrow$ find the values of parameters that give the best description to a set of measured values of y (data obtained from experiments or simulations).

Maximum likelihood

- Maximum likelihood $\Rightarrow$ cannonical ensembles in statistical physics.
- Let y_i be the measured value of y at $x = x_i$ in the set of data $\{(x_i, y_i)\}$.
- Due to the uncertainties or experimental or systematic errors and so on:
  - y_i is likely to be different from what we expected if we repeat the measurement
  - By the central limit theorem

$$p(y_i)dy_i = \frac{1}{\sigma_i\sqrt{2\pi}} e^{(y_i - f_i)^2/2\sigma_i^2} dy_i \quad (1)$$

where

$$f_i = f(a_1, a_2, \cdots, a_m; x_i)$$

and σ_i^2 is the variance.

Maximum likelihood

Def.: Likelihood Function

The likelihood function for $\{y_i\}$ is defined as

$$\begin{aligned}\mathcal{L}(a_1, a_2, \cdots, a_m) &= \prod_{i=1}^{N} p(y_i) \\ &= \prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}} e^{(y_i - f_i)^2/2\sigma_i^2} \\ &= \left\{\prod_{i=1}^{N} \frac{1}{\sigma_i\sqrt{2\pi}}\right\} \exp\left[-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{y_i - f_i}{\sigma_i}\right)^2\right] \qquad (2)\end{aligned}$$

Maximum likelihood

Def.: χ^2

$$\chi^2 \equiv \sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i} \right)^2 \tag{3}$$

Minimize χ^2 in the parameter space

$$\frac{\partial \chi^2}{\partial a_k} = -2 \sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i^2} \right) \frac{\partial f_i}{\partial a_k} = 0 \tag{4}$$

where $k = 1, 2, \cdots, m$.

Least-square Fit to a Straight Line

Let

$$a_1 = a, \qquad a_2 = b$$

Then

$$f(a, b; x) = a + bx$$

and

$$\frac{\partial f}{\partial a} = 1, \text{ and } \quad \frac{\partial f}{\partial b} = x$$

From Eq. (4)

$$\sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i^2} \right) = 0, \text{ and } \quad \sum_{i=1}^{N} \left(\frac{y_i - f_i}{\sigma_i^2} \right) x_i = 0$$

Least-square Fit to a Straight Line

Or using $f_i = a + bx_i$

$$a\sum_{i=1}^{N}\frac{1}{\sigma_i^2} + b\sum_{i=1}^{N}\frac{x_i}{\sigma_i^2} = \sum_{i=1}^{N}\frac{y_i}{\sigma_i^2}$$

$$a\sum_{i=1}^{N}\frac{1}{\sigma_i^2} + b\sum_{i=1}^{N}\frac{x_i^2}{\sigma_i^2} = \sum_{i=1}^{N}\frac{x_iy_i}{\sigma_i^2}$$

Least-square Fit to a Straight Line

In the matrix form

$$\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} = \begin{pmatrix} \theta \\ \phi \end{pmatrix}$$

where

$$\alpha = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}, \quad \beta = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2}, \quad \gamma = \beta$$

$$\delta = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2}, \quad \theta = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2}, \quad \phi = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2}$$

Least-square Fit to a Straight Line

What we have to calculate are

$$a = \frac{1}{D}\begin{vmatrix} \theta & \beta \\ \phi & \delta \end{vmatrix} = \frac{1}{D}(\theta\delta - \beta\phi)$$

$$b = \frac{1}{D}\begin{vmatrix} \alpha & \theta \\ \gamma & \phi \end{vmatrix} = \frac{1}{D}(\alpha\phi - \gamma\theta)$$

where D is the value of the determinant

$$D = \begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = \alpha\delta - \beta\gamma = \alpha\delta - \beta^2$$

Least-square Fit to a Straight Line

Errors: Uncertainties in the parameters

Error propagation:

$$\sigma_a^2 = \sum_{i=1}^{N} \left\{ \left(\frac{\partial a}{\partial y_i} \right)^2 \sigma_i^2 \right\} \quad \sigma_b^2 = \sum_{i=1}^{N} \left\{ \left(\frac{\partial b}{\partial y_i} \right)^2 \sigma_i^2 \right\} \tag{5}$$

From the calculation of a, b, and D

$$\frac{\partial a}{\partial y_i} = \frac{1}{D} \left(\frac{1}{\sigma_i^2} \delta - \frac{x_i}{\sigma_i^2} \beta \right), \qquad \frac{\partial b}{\partial y_i} = \frac{1}{D} \left(\frac{x_i}{\sigma_i^2} \alpha - \frac{1}{\sigma_i^2} \gamma \right) \tag{6}$$

Least-square Fit to a Straight Line

Errors: Uncertainties in the parameters

Inserting Eq. (6) into Eq. (5)

$$\sigma_a^2 = \frac{1}{D^2}\sum_{i=1}^{N}\left\{\sigma_i^2\left(\frac{1}{\sigma_i^2}\delta - \frac{x_i}{\sigma_i^2}\beta\right)^2\right\} = \frac{\delta}{D} \quad (7)$$

$$\sigma_b^2 = \frac{1}{D^2}\sum_{i=1}^{N}\left\{\sigma_i^2\left(\frac{x_i}{\sigma_i^2}\alpha - \frac{1}{\sigma_i^2}\gamma\right)^2\right\} = \frac{\alpha}{D} \quad (8)$$