Computational Astrophysics

E. Larrañaga

Observatorio Astronómico Nacional
Universidad Nacional de Colombia

April 30, 2019

Outline

Optimization

Empirical relationships (e.g., the $M - \sigma$ relation for galaxies) are typically established by taking experimental/observational data and fitting an analytic function to them.
In this section, we will introduce the most common curve fitting methods.

Curve Fitting Criteria

N data points (x_i, y_i)
Fit function $Y(x, \{a_j\})$
M parameters $\{a_j\}$
Least squares fit:

$$\Delta_i = Y(x_i, \{a_j\}) - y_i \ , \tag{1}$$

The goal is to minimize the function

$$\Delta(\{a_j\}) = \sum_{i=1}^{N} \Delta_i^2 = \sum_{i=1}^{N} \left(Y(x_i, \{a_j\}) - y_i\right)^2 \ . \tag{2}$$

The square of the difference is used because negative and positive variations would otherwise partially or fully cancel out, leading to a wrong result.

Curve Fitting Criteria

Observational data having an estimated error as $y_i \pm \sigma_i$

Chi-square function to minimize

$$\chi^2(\{a_j\}) = \sum_{i=1}^{N} \left(\frac{\Delta_i}{\sigma_i} \right)^2 = \sum_{i=1}^{N} \left(\frac{Y(x_i, \{a_j\}) - y_i}{\sigma_i} \right)^2 \tag{3}$$

Linear Regression

The simplest curve to fit some data is a stright line (*linear regression*).

$$Y(x, \{a_1, a_2\}) = a_1 + a_2 x \ , \tag{4}$$

The objective is to determine a_1 and a_2 such that

$$\chi^2(a_1, a_2) = \sum_{i=1}^{N} \frac{1}{\sigma_i^2}(a_1 + a_2 x_i - y_i)^2 \tag{5}$$

is minimized.

Linear Regression

Differentiating Eq. (5) and setting the result to zero:

$$\begin{aligned}
\frac{\partial \chi^2}{\partial a_1} &= 2\sum_{i=1}^{N} \frac{1}{\sigma_i^2}(a_1 + a_2 x_i - y_i) = 0 \ , \\
\frac{\partial \chi^2}{\partial a_2} &= 2\sum_{i=1}^{N} \frac{1}{\sigma_i^2}(a_1 + a_2 x_i - y_i) x_i = 0 \ .
\end{aligned} \tag{6}$$

Linear Regression

$$
\begin{aligned}
a_1 S + a_2 \Sigma x - \Sigma y = 0 \ , \\
a_1 \Sigma x + a_2 \Sigma x^2 - \Sigma xy = 0 \ ,
\end{aligned}
\tag{7}
$$

with

$$
\begin{gathered}
S = \sum_{i=1}^{N} \frac{1}{\sigma_i^2} \ , \quad \Sigma x = \sum_{i=1}^{N} \frac{x_i}{\sigma_i^2} \ , \quad \Sigma y = \sum_{i=1}^{N} \frac{y_i}{\sigma_i^2} \ , \\
\Sigma x^2 = \sum_{i=1}^{N} \frac{x_i^2}{\sigma_i^2} \ , \quad \Sigma xy = \sum_{i=1}^{N} \frac{x_i y_i}{\sigma_i^2} \ .
\end{gathered}
\tag{8}
$$

Linear Regression

Solving for the two unknowns a_1 and a_2 :

$$a_1 = \frac{\Sigma y \Sigma x^2 - \Sigma x \Sigma xy}{S\Sigma x^2 - (\Sigma x)^2} , \qquad a_2 = \frac{S\Sigma xy - \Sigma y \Sigma x}{S\Sigma x^2 - (\Sigma x)^2} . \tag{9}$$

- If all σ_i are identical, they will cancel out of the above equations and a_1 and a_2 will be independent of them.
- If the σ_i are unknown, then one can still use the χ^2 method and just sets $\sigma_i = 1$.

Linear Regression

Incorporating uncertainty in the x_i in the χ^2 fit must be handled by relating the error σ_i^x into an additional error in the y_i, σ_i^{extra}.

To first order, this can be done by writing

$$\sigma_{i,\text{extra}} = \left|\frac{\partial y}{\partial x}\right|_i \sigma_i^x \ , \tag{10}$$

where one needs an appropriate approximation for the slope $\partial y/\partial x$.
- If both σ_i and $\sigma_{i,\text{extra}}$ contribute significantly, one simply adds their squares: $\sigma_{i,\text{total}}^2 = \sigma_i^2 + \sigma_{i,\text{extra}}^2$.
- If the error in x_i or y_i is asymmetric about (x_i, y_i) one could weigh by the maximum of the left and right error, or use advanced techniques to incorporate this information.

Associated error bar $\sigma^2_{a_j}$ for the curve fit parameter a_j.
Using first-order error propagation, we have

$$\sigma^2_{a_j} = \sum_{i=1}^{N} \left(\frac{\partial a_j}{\partial y_i} \right)^2 \sigma_i^2 \ , \tag{11}$$

from which we obtain with Eq. (9)

$$\sigma_{a_1} = \sqrt{\frac{\Sigma x^2}{S \Sigma x^2 - (\Sigma x)^2}} \ , \qquad \sigma_{a_2} = \sqrt{\frac{S}{S \Sigma x^2 - (\Sigma x)^2}} \ . \tag{12}$$

Linear Regression

If the data set doesn't have an associated set of error bars, the error $\sigma_{a_j} = \sigma_0$ is estimated from the sample variance of the data,

$$\sigma_0^2 = \frac{1}{N-2} \sum_{i=1}^{N} \left(y_i - (a_1 + a_2 x_i)\right)^2 \quad . \tag{13}$$

The normalization factor $N - 2$ of the variance is due to the fact that we have taken out two parameters (a_1 and a_2) from the data.

Non-linear Fitting

Many non-linear fitting problems may be transformed to linear problems by a simple change of variables.

Example

Consider a power law

$$Z(t, \{\alpha, \beta\}) = \alpha t^{\beta} \ . \tag{14}$$

This may be rewritten as $Y(x, \{a_1, a_2\}) = a_1 + a_2 x$ with

$$Y = \log Z \ , \quad x = \log t \quad a_1 = \log \alpha \ , \quad a_2 = \beta \ . \tag{15}$$

Non-linear Fitting

Example

Consider an exponential

$$Z(t, \{\alpha, \beta\}) = \alpha e^{\beta x} \ . \tag{16}$$

This may be rewritten as $Y(x, \{a_1, a_2\}) = a_1 + a_2 x$ with

$$Y = \ln Z \ , \quad a_1 = \ln \alpha \ , \quad a_2 = \beta \ . \tag{17}$$

Next Class

Ordinary Differential Equations