Linear Least Squares Fitting

Bhas Bapat

IISER Pune

Nov 2014
What is Least Squares Fit?

- A procedure for finding the best-fitting curve to a given set of points by minimizing the sum of the squares of the offsets (called residuals) of the points from the curve.
- The sum of the squares of the offsets is used instead of the offset absolute values, to permit the residuals to be treated as a continuous differentiable quantity.
- However, this may cause outlying points to have a disproportionate effect on the fit.
What is Least Squares Fit?

- In practice, vertical offsets from a curve (or surface!) are minimized instead of perpendicular offsets.
- This provides a simpler analytic form for the fitting parameters and when noisy data points are few in number, the difference between vertical and perpendicular fits is quite small.
- Accommodates uncertainties of the data in $x$ and $y$
- The fitting technique can be easily generalized from a best-fit line to a best-fit polynomial when sums of vertical distances are used.
The linear least squares fitting technique is the simplest and most commonly applied form of linear regression (finding the best fitting straight line through a set of points.)

The fitting is **linear in the parameters to be determined**, it need not be linear in the independent variable $x$.

If the functional relationship between the two quantities being graphed is known, the data can often be transformed to obtain a straight line.

Some cases appropriate for a linear least squares fit:

$$v = u + at, \quad T \propto \sqrt{l}, \quad F = a/r^2, \quad V = U \exp(-t/\tau)$$
For non-linear least squares fitting to a number of unknown parameters, linear least squares fitting may be applied iteratively to a linearized form of the function until convergence is achieved.

Some examples where non-linear least squares fit is needed to determine the function parameters:

\[ \psi = A \sin(\omega t + \phi) \]

\[ f_N(x, \mu, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{1}{2} \left( \frac{x - \mu}{\sigma} \right)^2 \right] \]
The Procedure

- Find the residual (sum of the squares of the vertical deviations) of a set of data points from a function with \( m \) (linear) parameters

\[
R^2 = \sum_i [y_i - f(x_i, a_1, \ldots, a_m)]^2
\]

- The resulting residual is then minimized to find the best fit

\[
\frac{\partial (R^2)}{\partial a_i} = 0 \quad \forall \ i, \ (1 \ldots N)
\]

- A statistically more appropriate measure is

\[
\chi^2 = \frac{1}{N - m} \sum_i \frac{[y_i - f(x_i, a_1, \ldots, a_m)]^2}{\sigma_{y_i}^2 + \sum_j a_j^2 \sigma_{x_i}^2}
\]

- For a linear fit \( f(x, a, b) = a + bx \) with errors in \( y \)

\[
\chi^2 = \frac{1}{N - 2} \sum_i \frac{[y_i - a - bx_i]^2}{\sigma_{y_i}^2}
\]
The Procedure

So we get the equations

\[
\chi^2(a, b) = \sum \frac{[y_i - a - bx]^2}{\sigma^2_{y_i}}
\]

\[
\frac{\partial(\chi^2)}{\partial a} = -2 \sum \frac{[y_i - a - bx_i]}{\sigma^2_{y_i}} = 0
\]

\[
\frac{\partial(\chi^2)}{\partial b} = -2 \sum \frac{[y_i - a - bx_i]}{\sigma^2_{y_i}} x_i = 0
\]

Which leads to

\[
a \sum \frac{1}{\sigma^2_{y_i}} + b \sum \frac{x_i}{\sigma^2_{y_i}} = \sum \frac{y_i}{\sigma^2_{y_i}}
\]

\[
a \sum \frac{x_i}{\sigma^2_{y_i}} + b \sum \frac{x_i^2}{\sigma^2_{y_i}} = \sum \frac{x_i y_i}{\sigma^2_{y_i}}
\]
If $y$-errors are undefined, then we set them all to 1 (i.e. equal weights to all points), and the simultaneous equations for $a$ and $b$ are

\[
na + b \sum x_i = \sum y_i \\
a \sum x_i + b \sum x_i^2 = \sum x_i y_i
\]
The Fit Parameters

Define sums of squares:

\[ S_{xx} = \sum \frac{(x_i - \bar{x})^2}{\sigma^2_i} \]
\[ S_{yy} = \sum \frac{(y_i - \bar{y})^2}{\sigma^2_i} \]
\[ S_{xy} = \sum \frac{(x_i - \bar{x})(y_i - \bar{y})}{\sigma^2_i} \]

Which leads to

\[ b = \frac{S_{xy}}{S_{xx}} \quad \text{the regression coefficient} \quad \text{and} \quad a = \bar{y} - b\bar{x} \]

The quality of fit is parametrized by

\[ R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}} \quad \text{the correlation coefficient} \]
Fitting error

The fitting error at each point is

$$e_i = y_i - (a - bx_i)$$

The estimator of the variance of $e_i$ would be

$$s^2 = \sum \frac{e_i^2}{n-2}$$

or

$$s = \left[ \frac{S_{yy} - bS_{xy}}{n-2} \right]^{1/2}.$$  

The standard error in the fitted parameters is

$$SE(a) = s \left[ \frac{1}{n} + \frac{(\bar{x})^2}{S_{xx}} \right]^{1/2}$$

$$SE(b) = \frac{s}{S_{xx}^{1/2}}.$$
Examples

- **Physical Pendulum**

\[ T(\theta_0) = T_0 \left( 1 + \frac{1}{4} \sin^2 \frac{\theta_0}{2} + \cdots \right) \]

Plot \( T \) versus \( \frac{1}{4} \sin^2 \frac{\theta_0}{2} \)
or \( T \) versus \( \sin \theta_0 \)

- **Young's modulus**

\[ Y = \frac{MgL^3}{4BW^3\delta} \]

Plot \( \delta \) versus \( M \)
Examples

- Euler’s Method (Friction)
  \[ T_1 = T_2 \exp(\mu \theta) \]

  Plot \( T_1 \ vs \ exp(\mu \theta) \)
  or \( ln \ T_1 \ vs \ \mu \theta \) or \( ln \ T_1 \ vs \ ln \ T_2 \)

- Viscosity
  \[ \eta = \frac{2a^2 (\rho_{sphere} - \rho_{fluid}) g}{9VT} \]

  Plot \( VT \ vs \ a^2 (\rho_{sphere} - \rho_{fluid}) \)
Examples

- Helmholtz Coil
  \[ B = \mu_0 NI \frac{a^2}{2} \frac{1}{(z^2 + a^2)^{3/2}} \]

  Plot \( B \ versus \ z \) or \( B \ versus \ (z^2 + a^2)^{3/2} \)

- Faraday’s Law
  Plot \( V \ versus \ \theta_{\text{max}} \)

- Galvanometer constant
  Plot \( \theta \ versus \ 1/R \)
Examples

- Magnet Repulsion

\[ F = \left( \frac{\pi \mu_0}{4} \right) M^2 R^4 \left[ \frac{1}{z^2} + \frac{1}{(z+2h)^2} - \frac{2}{(z+h)^2} \right] \]

Plot \( mg \) versus \( z \) or \( mg \) versus \( \frac{1}{z^2} + \frac{1}{(z+2h)^2} - \frac{2}{(z+h)^2} \)

- Intensity of light

\[ I = \frac{a}{r^2} + c \]

Plot \( I \) versus \( 1/r^2 \) or \( I \) versus \( r \) or Plot \( \sqrt{I} \) versus \( 1/r \)
Inference on derived quantities

- The standard error on the linear fit parameters $a$ and $b$ have a bearing on the physical quantity (which is the aim of the experiment) that you derive based on the fit.
- Use standard error propagation method to find the error in the final quantity.
- As an example, in the Young’s Modulus experiment, the slope $b$ of the best fit line to the graph of $\delta$ versus $M$ is related to $Y$ through

$$Y = \frac{gL^3}{4BW^3} \frac{1}{b}$$

- The error in the fitted value of $b$ propagates to an error in $Y$:

$$Y = \frac{gL^3}{4BW^3} \left[ \frac{1}{b} \pm \frac{-SE(b)}{b^2} \right]$$
A few points are extremely important in this context

- There has to be a physical basis for choosing a function.
- A good fit (small chi-square) for a particular function does not imply a cause–effect relationship or the correctness of the function.
- A small chi-square alone is not adequate. The errors in fitted parameters should also be small.
- Ensure that there is a large number of data points for fitting. Linear least squares depends strongly on the problem being overdetermined.
- **Rule of thumb:** the number of points should be at least thrice the number of parameters to be fitted.