Semester-VI

B.Sc (Honours) in Physics

DSE 4: Experimental Techniques

Lecture on Curve fitting

Discussed by Dr. K R Sahu

Lecture- Vb

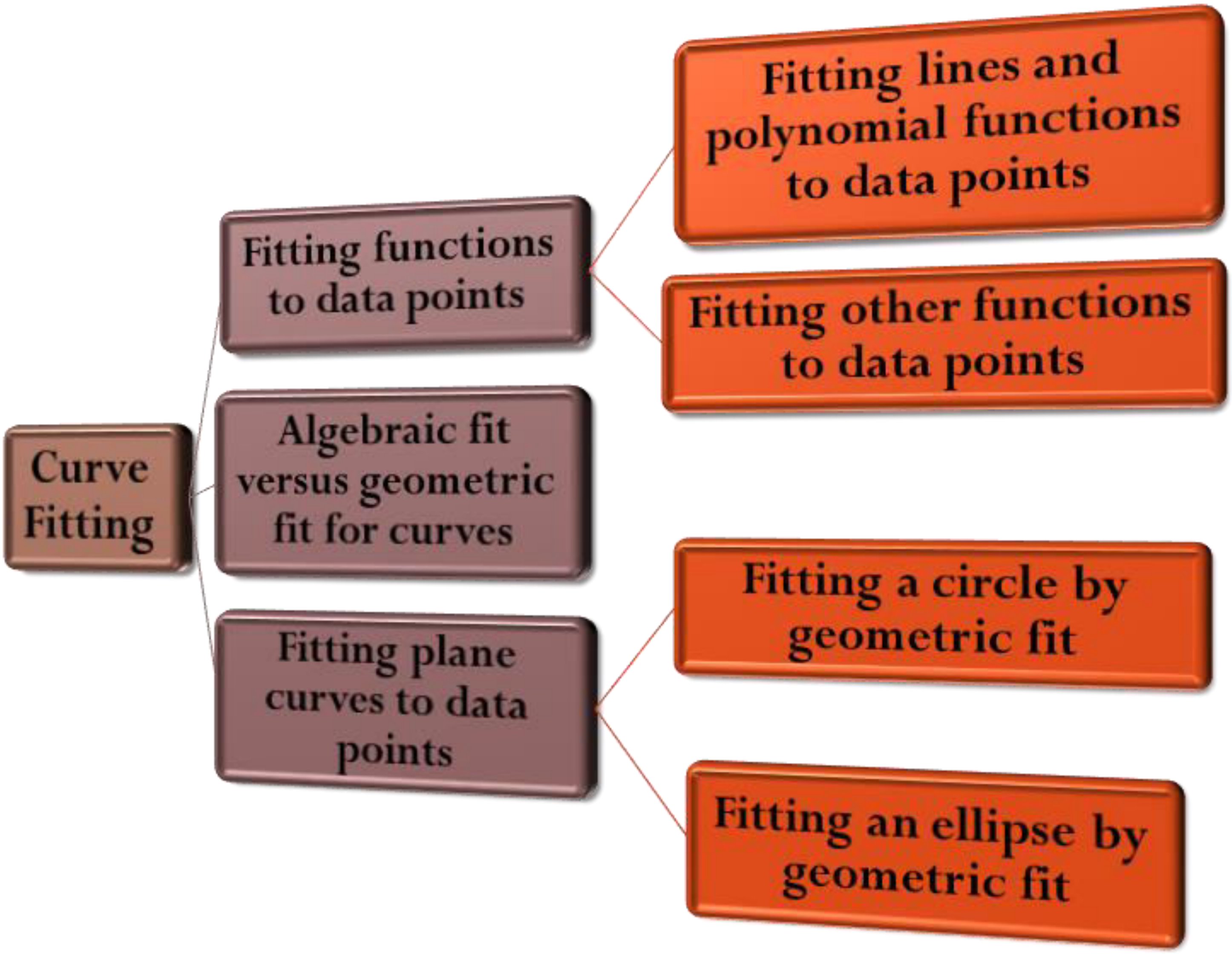
Syllabus

- ❑ **Measurements**
  - ❑ Accuracy and precision and Significant figures.
  - ❑ Error and uncertainty analysis.
  - ❑ Types of errors:
    - ❑ Gross error,
    - ❑ Systematic error,
    - ❑ Random error.
  - ❑ Statistical analysis of data
    - ❑ Arithmetic mean,
    - ❑ Deviation from mean,
    - ❑ Average deviation,
    - ❑ Standard deviation,
    - ❑ Chi-square and
  - ❑ **Curve fitting.**
  - ❑ Guassian distribution.

Curve fitting

Curve fitting is the process of constructing a curve, or mathematical function, that has the ***best fit to a series of data points***, possibly subject to constraints. Curve fitting can involve either interpolation, where an exact fit to the data is required, or smoothing, in which a "**smooth**" function is constructed that approximately fits the data. A related topic is regression analysis, which focuses more on questions of statistical inference such as how much uncertainty is present in a curve that is fit to data observed with random errors. Fitted curves can be used as an aid for data visualization, to infer values of a function where no data are available, and to summarize the relationships among two or more variables. Extrapolation refers to the use of a fitted curve beyond the range of the observed data, and is subject to a degree of uncertainty since it may reflect the method used to construct the curve as much as it reflects the observed data.

Types of Curve fitting



Mathematical Calculation of Curve fitting and with Examples

Least square regression

Given data for discrete values, derive a single curve that represents the general trend of the data. When the given data exhibit a significant degree of error or noise.

Simple Linear Regression

Fitting a straight line to a set of paired observations $(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)$.
Mathematical expression for the straight line (model)

$$y = a_0 + a_1 x$$

where a_0 is the intercept, and a_1 is the slope.
Define

$$e_i = y_{i,measured} - y_{i,model} = y_i - (a_0 + a_1 x_i)$$

Criterion for a best fit:

$$\min S_r = \min_{a_0,a_1} \sum_{i=1}^{n} e_i^2 = \min_{a_0,a_1} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Find a_0 and a_1:

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i) = 0 \quad (1)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} [(y_i - a_0 - a_1 x_i) x_i] = 0 \quad (2)$$

From (1), $\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1 x_i = 0$, or

$$na_0 + \sum_{i=1}^{n} x_i a_1 = \sum_{i=1}^{n} y_i \quad (3)$$

From (2), $\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} a_0 x_i - \sum_{i=1}^{n} a_1 x_i^2 = 0$, or

$$\sum_{i=1}^{n} x_i a_0 + \sum_{i=1}^{n} x_i^2 a_1 = \sum_{i=1}^{n} x_i y_i \quad (4)$$

(3) and (4) are called normal equations.

From (3),

$$a_0 = \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i a_1 = \bar{y} - \bar{x} a_1$$

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$.

From (4), $\sum_{i=1}^{n} x_i(\frac{1}{n}\sum_{i=1}^{n} y_i - \frac{1}{n}\sum_{i=1}^{n} x_i a_1) + \sum_{i=1}^{n} x_i^2 a_1 = \sum_{i=1}^{n} x_i y_i$,

$$a_1 = \frac{\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2}$$

or

$$a_1 = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}$$

Definitions:

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

Standard error of the estimate:

$$S_{y/x} = \sqrt{\frac{S_r}{n-2}}$$

— Spread around the regression line

Standard deviation of data points

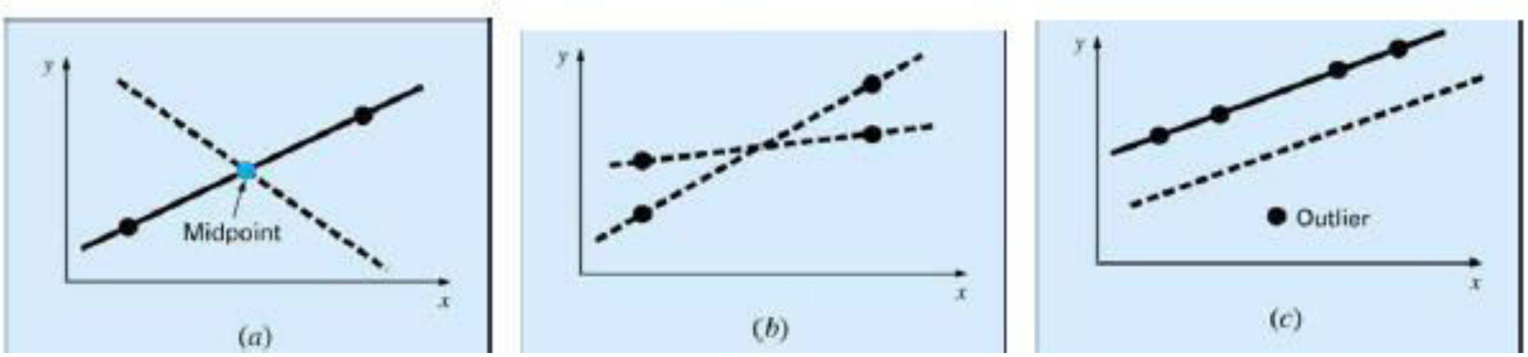
$$S_y = \sqrt{\frac{S_t}{n-1}} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}}$$

where $S_t = \sum_{i=1}^{n}(y_i - \bar{y})^2$.
— Spread around the mean value $\bar{y}$.

Correlation coefficient:

$$r = \sqrt{\frac{S_t - S_r}{S_t}}$$

— Improvement or error reduction due to describing the data in terms of a straight line rather than as an average value.



Other criteria for regression (a) min ,$\sum \mathbf{e_i}$ (b) min $\sum |\mathbf{e_i}|$ and (c) min max $\sum \mathbf{e_i}$

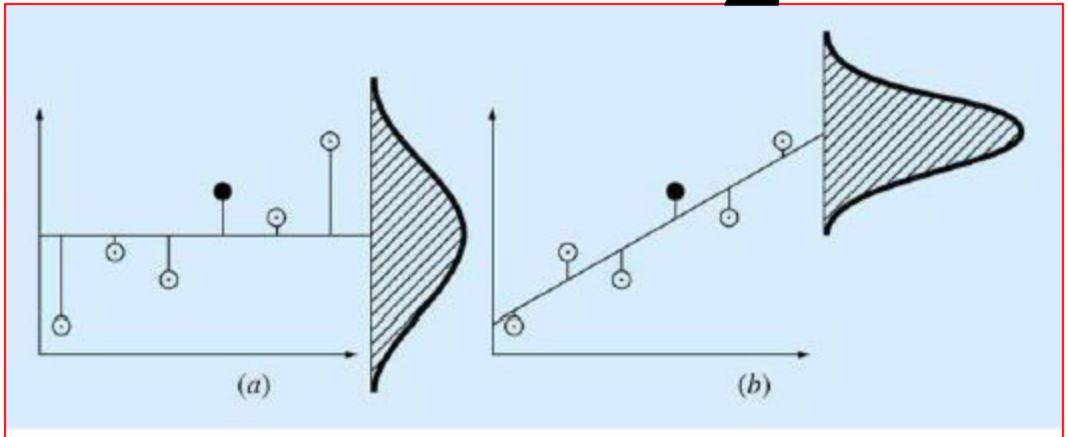


Figure 1: (a) Spread of data around mean of dependent variable, (b) spread of data around the best-fit line

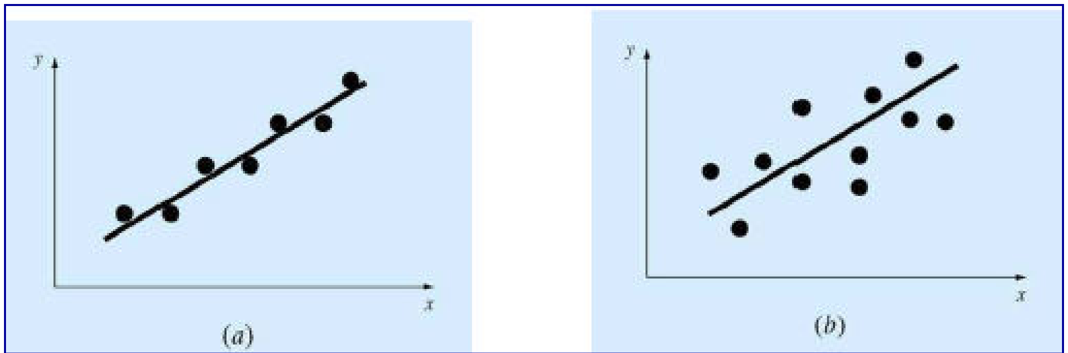


Illustration of linear regression with (a) small and (b) large residual errors

Example:

x	1	2	3	4	5	6	7
y	0.5	2.5	2.0	4.0	3.5	6.0	5.5

$\sum x_i = 1 + 2 + \ldots + 7 = 28$
$\sum y_i = 0.5 + 2.5 + \ldots + 5.5 = 24$
$\sum x_i^2 = 1^2 + 2^2 + \ldots + 7^2 = 140$
$\sum x_i y_i = 1 \times 0.5 + 2 \times 2.5 + \ldots + 7 \times 5.5 = 119.5$
$a_1 = \frac{n\sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} = \frac{7\times119.5-28\times24}{7\times140-28^2} = 0.8393$
$a_0 = \bar{y} - \bar{x}a_1 = \frac{1}{n}\sum y_i - a_1\frac{1}{n}\sum x_i = \frac{1}{7}\times 24 - 0.8393 \times \frac{1}{7} \times 28 = 0.07143.$
Model: $y = 0.07143 + 0.8393x.$

$S_r = \sum_{i=1}^n e_i^2, e_i = y_i - (a_0 + a_1 x_i)$
$e_1 = 0.5 - 0.07143 - 0.8393 \times 1 = -0.410$
$e_2 = 2.5 - 0.07143 - 0.8393 \times 2 = 0.750$
$e_3 = 2.0 - 0.07143 - 0.8393 \times 3 = -0.589$
$\ldots$
$e_7 = 5.5 - 0.07143 - 0.8393 \times 7 = -0.446$
$S_r = (-0.410)^2 + 0.750^2 + (-0.589)^2 + \ldots + 0.446^2 = 2.9911$

$S_t = \sum_{i=1}^n (y_i - \bar{y})^2 = 22.714$
Standard deviation of data points:
$S_y = \sqrt{\frac{S_t}{n-1}} = \sqrt{\frac{22.714}{6}} = 1.946$
Standard error of the estimate:
$S_{y/x} = \sqrt{\frac{S_r}{n-2}} = \sqrt{\frac{2.9911}{7-2}} = 0.774$
$S_{y/x} < S_y$, $S_r < S_t$.

Correlation coefficient $r = \sqrt{\frac{S_t - S_r}{S_t}} = \sqrt{\frac{22.714-2.9911}{22.714}} = 0.932$

Polynomial Regression

Given data (x_i, y_i), $i = 1, 2, \ldots, n$, fit a second order polynomial

$$y = a_0 + a_1 x + a_2 x^2$$

$e_i = y_{i,measured} - y_{i,model} = y_i - (a_0 + a_1 x_i + a_2 x_i^2)$

Criterion for a best fit:

$$\min S_r = \min_{a_0,a_1,a_2} \sum_{i=1}^{n} e_i^2 = \min_{a_0,a_1,a_2} \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i - a_2 x_i^2)^2$$

Find a_0, a_1, and a_2:

$$\frac{\partial S_r}{\partial a_0} = -2\sum_{i=1}^{n}(y_i - a_0 - a_1x_i - a_2x_i^2) = 0 \quad (1)$$

$$\frac{\partial S_r}{\partial a_1} = -2\sum_{i=1}^{n}[(y_i - a_0 - a_1x_i - a_2x_i^2)x_i] = 0 \quad (2)$$

$$\frac{\partial S_r}{\partial a_2} = -2\sum_{i=1}^{n}[(y_i - a_0 - a_1x_i - a_2x_i^2)x_i^2] = 0 \quad (3)$$

From (1), $\sum_{i=1}^{n} y_i - \sum_{i=1}^{n} a_0 - \sum_{i=1}^{n} a_1x_i - \sum_{i=1}^{n} a_2x_i^2 = 0$, or

$$na_0 + \sum_{i=1}^{n} x_ia_1 + \sum_{i=1}^{n} x_i^2a_2 = \sum_{i=1}^{n} y_i \quad (1')$$

From (2), $\sum_{i=1}^{n} x_iy_i - \sum_{i=1}^{n} a_0x_i - \sum_{i=1}^{n} a_1x_i^2 - \sum_{i=1}^{n} x_i^3a_2 = 0$, or

$$\sum_{i=1}^{n} x_ia_0 + \sum_{i=1}^{n} x_i^2a_1 + \sum_{i=1}^{n} x_i^3a_2 = \sum_{i=1}^{n} x_iy_i \quad (2')$$

From (3), $\sum_{i=1}^{n} x_i^2y_i - \sum_{i=1}^{n} a_0x_i^2 - \sum_{i=1}^{n} a_1x_i^3 - \sum_{i=1}^{n} x_i^4a_2 = 0$, or

$$\sum_{i=1}^{n} x_i^2a_0 + \sum_{i=1}^{n} x_i^3a_1 + \sum_{i=1}^{n} x_i^4a_2 = \sum_{i=1}^{n} x_i^2y_i \quad (3')$$

Comments:

- The problem of determining a least-squares second order polynomial is equivalent to solving a system of 3 simultaneous linear equations.
- In general, to fit an m-th order polynomial

$$y = a_0 + a_1 x_1 + a_2 x^2 + \ldots + a_m x^m$$

using least-square regression is equivalent to solving a system of $(m+1)$ simultaneous linear equations.

Standard error: $S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$

Multiple Linear Regression

Multiple linear regression is used when y is a linear function of 2 or more independent variables.

Model: $y = a_0 + a_1 x_1 + a_2 x_2$.

Given data (x_{1i}, x_{2i}, y_i), $i = 1, 2, \ldots, n$

$e_i = y_{i,measured} - y_{i,model}$

$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i})^2$

Find a_0, a_1, and a_2 to minimize S_r.

$$\frac{\partial S_r}{\partial a_0} = -2 \sum_{i=1}^{n} (y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) = 0 \quad (1)$$

$$\frac{\partial S_r}{\partial a_1} = -2 \sum_{i=1}^{n} [(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) x_{1i}] = 0 \quad (2)$$

$$\frac{\partial S_r}{\partial a_2} = -2 \sum_{i=1}^{n} [(y_i - a_0 - a_1 x_{1i} - a_2 x_{2i}) x_{2i}] = 0 \quad (3)$$

From (1), $n a_0 + \sum_{i=1}^{n} x_{1i} a_1 + \sum_{i=1}^{n} x_{2i} a_2 = \sum_{i=1}^{n} y_i \quad (1')$

From (2), $\sum_{i=1}^{n} x_{1i} a_0 + \sum_{i=1}^{n} x_{1i}^2 a_1 + \sum_{i=1}^{n} x_{1i} x_{2i} a_2 = \sum_{i=1}^{n} x_{1i} y_i \quad (2')$

From (3), $\sum_{i=1}^{n} x_{2i}a_0 + \sum_{i=1}^{n} x_{1i}x_{2i}a_1 + \sum_{i=1}^{n} x_{2i}^2 a_2 = \sum_{i=1}^{n} x_{2i}y_i \quad (3')$

$$\begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} \\ \sum x_{1i} & \sum x_{1i}^2 & \sum x_{1i}x_{2i} \\ \sum x_{2i} & \sum x_{1i}x_{2i} & \sum x_{2i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{bmatrix}$$

Standard error: $S_{y/x} = \sqrt{\frac{S_r}{n-(m+1)}}$

General Linear Least Squares

Model:

$$y = a_0 Z_0 + a_1 Z_1 + a_2 Z_2 + \ldots + a_m Z_m$$

where $Z_0, Z_1, \ldots, Z_m$ are $(m+1)$ different functions.

Special cases:

- Simple linear LSR: $Z_0 = 1$, $Z_1 = x$, $Z_i = 0$ for $i \geq 2$
- Polynomial LSR: $Z_i = x^i$ ($Z_0 = 1$, $Z_1 = x$, $Z_2 = x^2$, ...)
- Multiple linear LSR: $Z_0 = 1$, $Z_i = x_i$ for $i \geq 1$

"Linear" indicates the model's dependence on its parameters, a_i's. The functions can be highly non-linear.

$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_{i,measured} - y_{i,model})^2$

Given data $(Z_{0i}, Z_{1i}, \ldots, Z_{mi}, y_i)$, $i = 1, 2, \ldots, n$,

$$S_r = \sum_{i=1}^{n} (y_i - \sum_{j=0}^{m} a_j Z_{ji})^2$$

Find a_j, $j = 0, 1, 2, \ldots, m$ to minimize S_r.

$$\frac{\partial S_r}{\partial a_k} = -2 \sum_{i=1}^{n} (y_i - \sum_{j=0}^{m} a_j Z_{ji}) \cdot Z_{ki} = 0$$

$$\sum_{i=1}^{n} y_i Z_{ki} = \sum_{i=1}^{n} \sum_{j=0}^{m} Z_{ki} Z_{ji} a_j, \quad k = 0, 1, \ldots, m$$

$$\sum_{j=0}^{m} \sum_{i=1}^{n} Z_{ki} Z_{ji} a_j = \sum_{i=1}^{n} y_i Z_{ki}$$

$$Z^T Z A = Z^T Y$$

where

$$Z = \begin{bmatrix} Z_{01} & Z_{11} & \ldots & Z_{m1} \\ Z_{02} & Z_{12} & \ldots & Z_{m2} \\ & \ldots & & \\ Z_{0n} & Z_{1n} & \ldots & Z_{mn} \end{bmatrix}$$

Polynomial Interpolation

Given (n + 1) data points, (x_i, y_i), i = 0, 1, 2, . . . , n, there is one and only one polynomial of order n that passes through all the points.

Newton's Divided-Difference Interpolating Polynomials

Linear Interpolation

Given (x_0, y_0) and (x_1, y_1)

$$\frac{y_1 - y_0}{x_1 - x_0} = \frac{f_1(x) - y_0}{x - x_0}$$

$$f_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0)$$

$f_1(x)$: first order interpolation

A smaller interval, i.e., $|x_1 - x_0|$ closer to zero, leads to better approximation.

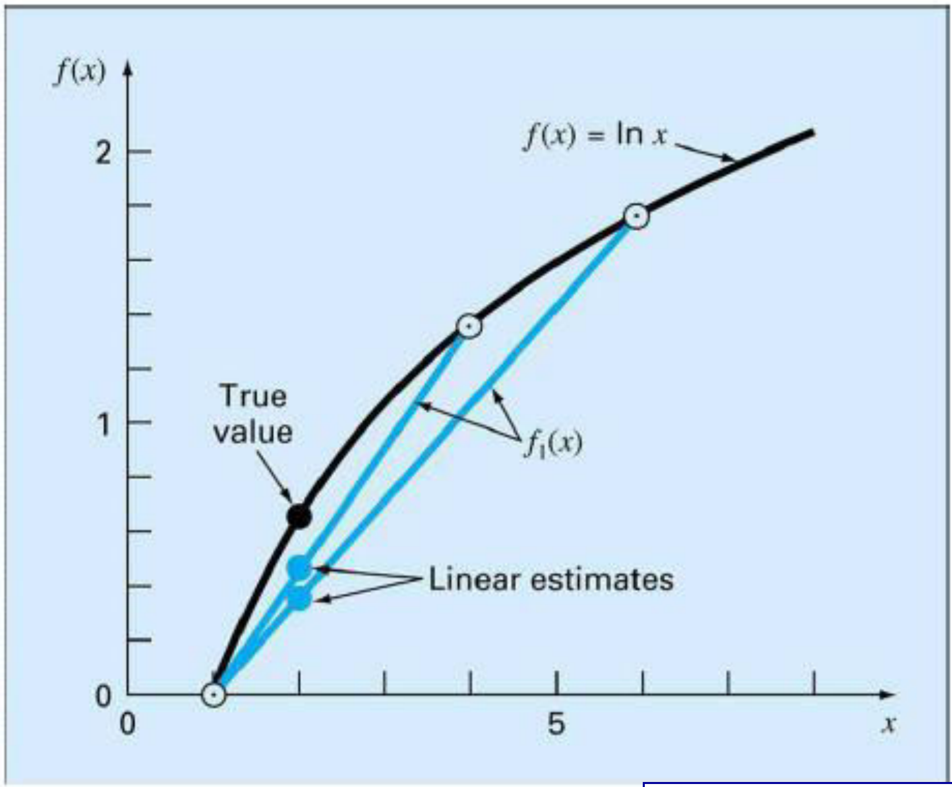
Example: Given $\ln 1 = 0$, $\ln 6 = 1.791759$, use linear interpolation to find $\ln 2$.

Solution:

$f_1(2) = \ln 2 = \ln 1 + \frac{\ln 6 - \ln 1}{6-1} \times (2-1) = 0.3583519$

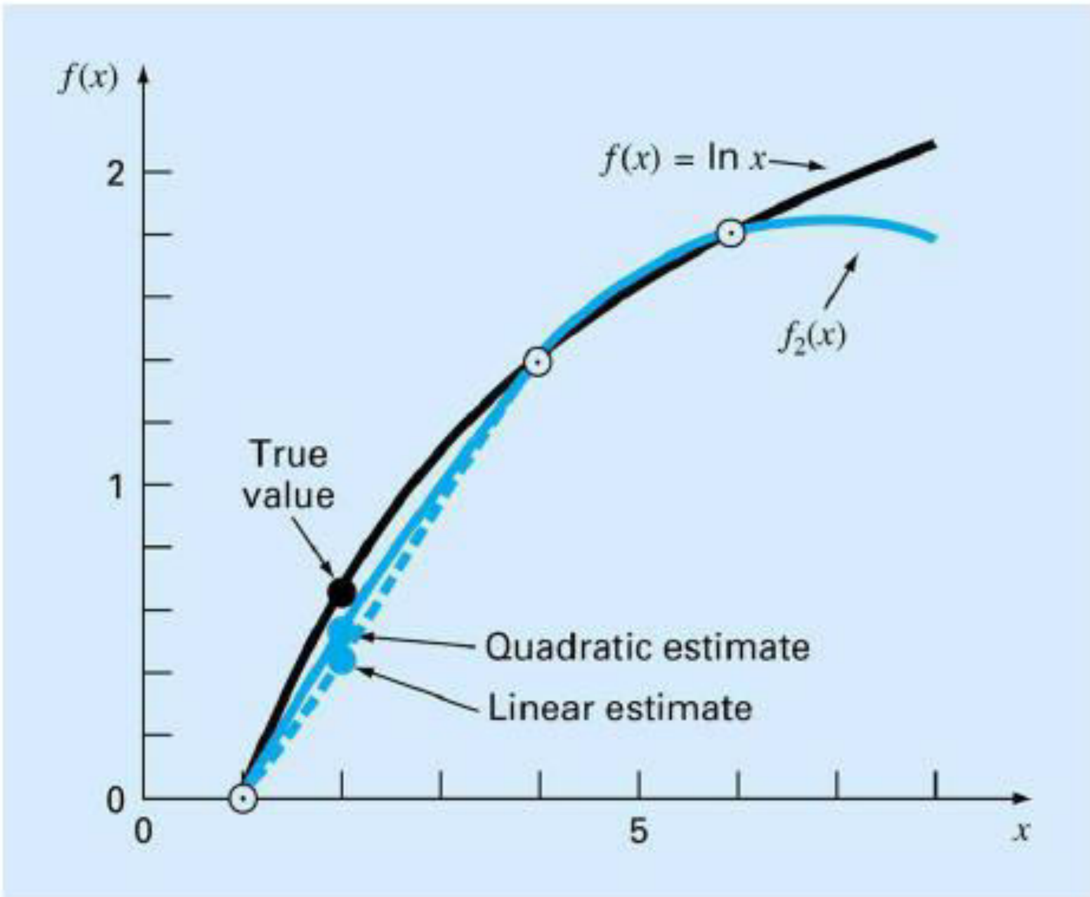
True solution: $\ln 2 = 0.6931472$.

$\epsilon_t = |\frac{f_1(2) - \ln 2}{\ln 2}| \times 100\% = |\frac{0.3583519 - 0.6931472}{0.6931472}| \times 100\% = 48.3\%$



A smaller interval provides a better estimate

Quadratic interpolation provides a better estimate than linear interpolation



Quadratic Interpolation

Given 3 data points, (x_0, y_0), (x_1, y_1), and (x_2, y_2), we can have a second order polynomial

$f_2(x) = b_0 + b_1(x - x_0) + b_2(x - x_0)(x - x_1)$
$f_2(x_0) = b_0 = y_0$
$f_2(x_1) = b_0 + b_1(x_1 - x_0) = y_1, \rightarrow \; b_1 = \frac{y_1 - y_0}{x_1 - x_0}$
$f_2(x_2) = b_0 + b_1(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1) = y_2, \rightarrow \; b_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0}$ (*)

Proof (*):

$$
\begin{aligned}
b_2 &= \frac{y_2 - b_0 - b_1(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)} = \frac{y_2 - y_0 - \frac{(y_1 - y_0)(x_2 - x_0)}{x_1 - x_0}}{(x_2 - x_0)(x_2 - x_1)} \\
&= \frac{(y_2 - y_0)(x_1 - x_0) - (y_1 - y_0)(x_2 - x_0)}{(x_2 - x_0)(x_2 - x_1)(x_1 - x_0)} \\
&= \frac{y_2(x_1 - x_0) - y_0 x_1 + y_0 x_0 - (y_1 - y_0)x_2 + y_1 x_0 - y_0 x_0}{(x_2 - x_0)(x_2 - x_1)(x_1 - x_0)} \\
&= \frac{y_2(x_1 - x_0) - y_1 x_1 + y_1 x_0 - (y_1 - y_0)x_2 + y_1 x_1 - y_0 x_1}{(x_2 - x_0)(x_2 - x_1)(x_1 - x_0)} \\
&= \frac{(y_2 - y_1)(x_1 - x_0) - (y_1 - y_0)(x_2 - x_1)}{(x_2 - x_0)(x_2 - x_1)(x_1 - x_0)}
\end{aligned}
$$

Comments: In the expression of $f_2(x)$,

- $b_0 + b_1(x - x_0)$ is linear interpolating from (x_0, y_0) and (x_1, y_1), and
- $+b_2(x - x_0)(x - x_1)$ introduces second order curvature.

Example: Given $\ln 1 = 0$, $\ln 4 = 1.386294$, and $\ln 6 = 1.791759$, find $\ln 2$.

Solution:

$(x_0, y_0) = (1, 0)$, $(x_1, y_1) = (4, 1.386294)$, $(x_2, y_2) = (6, 1.791759)$

$b_0 = y_0 = 0$

$b_1 = \frac{y_1 - y_0}{x_1 - x_0} = \frac{1.386294 - 0}{4 - 1} = 0.4620981$

$b_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0} = \frac{\frac{1.791759 - 1.386294}{6 - 4} - 0.4620981}{6 - 1} = -0.0518731$

$f_2(x) = 0.4620981(x - 1) - 0.0518731(x - 1)(x - 4)$

$f_2(2) = 0.565844$

$\epsilon_t = |\frac{f_2(2) - \ln 2}{\ln 2}| \times 100\% = 18.4\%$

Straightforward Approach

$$y = a_0 + a_1 x + a_2 x^2$$

or

$$\begin{aligned} a_0 + a_1 x_0 + a_2 x_0^2 &= y_0 \\ a_0 + a_1 x_1 + a_2 x_1^2 &= y_1 \\ a_0 + a_1 x_2 + a_2 x_2^2 &= y_2 \end{aligned}$$

$$\begin{bmatrix} 1 & x_0 & x_0^2 \\ 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix}$$

General Form of Newton's Interpolating Polynomial

Given $(n+1)$ data points, (x_i, y_i), $i = 0, 1, \ldots, n$, fit an n-th order polynomial

$$f_n(x) = b_0 + b_1(x - x_0) + \ldots + b_n(x - x_0)(x - x_1)\ldots(x - x_n) = \sum_{i=0}^{n} b_i \prod_{j=0}^{i-1}(x - x_j)$$

find $b_0, b_1, \ldots, b_n$.

$x = x_0$, $y_0 = b_0$ or $b_0 = y_0$.

$x = x_1$, $y_1 = b_0 + b_1(x_1 - x_0)$, then $b_1 = \frac{y_1 - y_0}{x_1 - x_0}$

Define $b_1 = f[x_1, x_0] = \frac{y_1 - y_0}{x_1 - x_0}$.

$x = x_2$, $y_2 = b_0 + b_1(x_2 - x_0) + b_2(x_2 - x_0)(x_2 - x_1)$, then $b_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0}$

Define $f[x_2, x_1, x_0] = \frac{f[x_2, x_1] - f[x_1, x_0]}{x_2 - x_0}$, then $b_2 = f[x_2, x_1, x_0]$.

$\ldots$

$x = x_n$, $b_n = f[x_n, x_{n-1}, \ldots, x_1, x_0] = \frac{f[x_n, x_{n-1}, \ldots, x_1] - f[x_{n-1}, \ldots, x_1, x_0]}{x_n - x_0}$

6 Lagrange Interpolating Polynomials

The Lagrange interpolating polynomial is a reformulation of the Newton's interpolating polynomial that avoids the computation of divided differences. The basic format is

$$f_n(x) = \sum_{i=0}^{n} L_i(x) f(x_i)$$

where $L_i(x) = \prod_{j=0, j\neq i}^{n} \frac{x-x_j}{x_i-x_j}$

Linear Interpolation ($n = 1$)

$f_1(x) = \sum_{i=0}^{1} L_i(x) f(x_i) = L_0(x)y_0 + L_1(x)y_1 = \frac{x-x_1}{x_0-x_1}y_0 + \frac{x-x_0}{x_1-x_0}y_1$

$(f_1(x) = y_0 + \frac{y_1-y_0}{x_1-x_0}(x - x_0))$

Second Order Interpolation ($n = 2$)

$f_2(x) = \sum_{i=0}^{2} L_i(x) f(x_i) = L_0(x)y_0 + L_1(x)y_1 + L_2(x)y_2 = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)}y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)}y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)}y_2$

Example: Given $\ln 1 = 0$, $\ln 4 = 1.386294$, and $\ln 6 = 1.791759$, find $\ln 2$.

Solution:

$(x_0, y_0) = (1, 0)$, $(x_1, y_1) = (4, 1.386294)$, $(x_2, y_2) = (6, 1.791759)$

$f_1(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = \frac{x-4}{1-4} \times 0 + \frac{x-1}{4-1} \times 1.386294 = 0.4620981$

$f_2(x) = \frac{(x-x_1)(x-x_2)}{(x_0-x_1)(x_0-x_2)} y_0 + \frac{(x-x_0)(x-x_2)}{(x_1-x_0)(x_1-x_2)} y_1 + \frac{(x-x_0)(x-x_1)}{(x_2-x_0)(x_2-x_1)} y_2 = \frac{(x-4)(x-6)}{(1-4)(1-6)} \times 0 + \frac{(x-1)(x-6)}{(4-1)(4-6)} \times 1.386294 + \frac{(x-1)(x-4)}{(6-1)(6-4)} \times 1.791760 = 0.565844$

Example: Find $f(2.6)$ by interpolating the following table of values.

i	x_i	y_i
1	1	2.7183
2	2	7.3891
3	3	20.0855

(1) Use Lagrange interpolation

$f_2(x) = \sum_{i=1}^{3} L_i(x) f(x_i)$, $L_i(x) = \prod_{j=1, j\neq i}^{3} \frac{x - x_j}{x_i - x_j}$

$L_1(x) = \frac{(x-x_2)(x-x_3)}{(x_1-x_2)(x_1-x_3)} = \frac{(2.6-2)(2.6-3)}{(1-2)(1-3)} = -0.12$

$L_2(x) = \frac{(x-x_1)(x-x_3)}{(x_2-x_1)(x_2-x_3)} = \frac{(2.6-1)(2.6-3)}{(2-1)(2-3)} = 0.64$

$L_3(x) = \frac{(x-x_1)(x-x_2)}{(x_3-x_1)(x_3-x_2)} = \frac{(2.6-1)(2.6-2)}{(3-1)(3-2)} = 0.48$

$f_2(2.6) = -0.12 \times 2.7183 + 0.64 \times 7.3891 + 0.48 \times 20.08853 = 14.0439$

(2) use Newton's interpolation

$f_2(x) = b_0 + b_1(x - x_1) + b_2(x - x_1)(x - x_2)$

$b_0 = y_1 = 2.7183$

$b_1 = \frac{y_2 - y_1}{x_2 - x_1} = \frac{7.3891 - 2.7183}{2 - 1} = 4.6708$

$b_2 = \frac{\frac{y_2 - y_1}{x_2 - x_1} - \frac{y_1 - y_0}{x_1 - x_0}}{x_2 - x_0} = \frac{\frac{20.0855 - 7.3891}{3 - 2} - 4.6708}{3 - 1} = 4.0128$

$f_2(2.6) = 2.7183 + 4.6708 \times (2.6 - 1) + 4.0128 \times (2.6 - 1)(2.6 - 2) = 14.0439$

(3) Use the straightforward method

$f_2(x) = a_0 + a_1 x + a_2 x^2$

$$a_0 + a_1 + a_2 \times 1^2 = 2.7183$$
$$a_0 + a_1 + a_2 \times 2^2 = 7.3891$$
$$a_0 + a_1 + a_2 \times 3^2 = 20.0855$$

or

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.7183 \\ 7.3891 \\ 20.0855 \end{bmatrix}$$

$[a_0\ a_1\ a_2]' = [6.0732; -7.3678\ 4.0129]'$

$f(2.6) = 6.0732 - 7.3678 \times 2.6 + 4.01219 \times 2.6^2 = 14.044.$

Example:

x_i	1	2	3	4
y_i	3.6	5.2	6.8	8.8

Model: $y = ax^b e^{cx}$

$\ln y = \ln a + b \ln x + cx$. Let $Y = \ln y$, $a_0 = \ln a$, $a_1 = b$, $x_1 = \ln x$, $a_2 = c$, and $x_2 = x$, then we have $Y = a_0 + a_1 x_1 + a_2 x_2$.

$x_{1,i}$	0	0.6931	1.0986	1.3863
$x_{2,i}$	1	2	3	4
Y_i	1.2809	1.6487	1.9169	2.1748

$\sum x_{1,i} = 3.1781$, $\sum x_{2,i} = 10$, $\sum x_{1,i}^2 = 3.6092$, $\sum x_{2,i}^2 = 30$, $\sum x_{1,i} x_{2,i} = 10.2273$, $\sum Y_i = 7.0213$, $\sum x_{1,i} Y_i = 6.2636$, $\sum x_{2,i} Y_i = 19.0280$. $n = 4$.

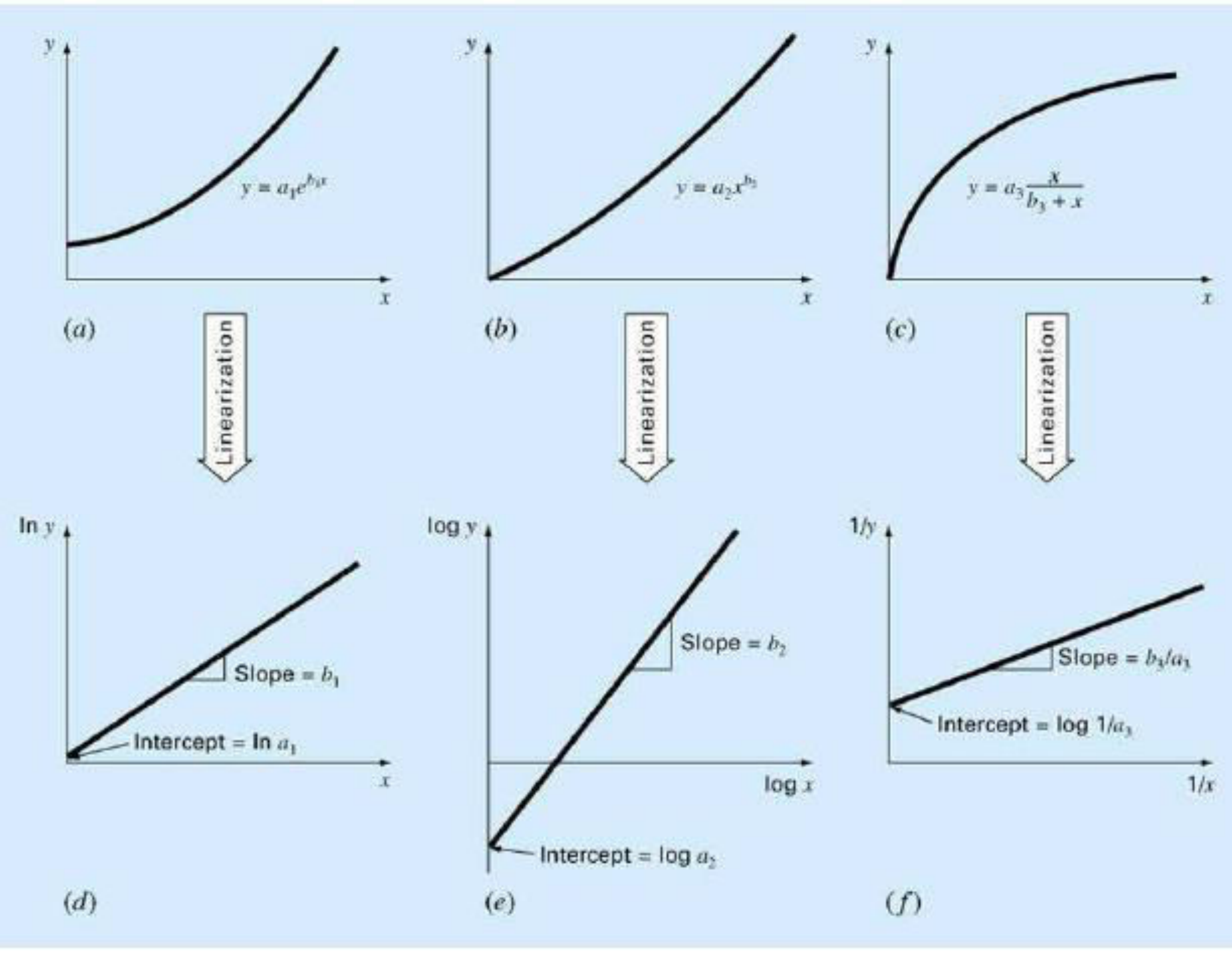
$$\begin{bmatrix} 1 & \sum x_{1,i} & \sum x_{2,i} \\ \sum x_{1,i} & \sum x_{1,i}^2 & \sum x_{2,i} x_{1,i} \\ \sum x_{2,i} & \sum x_{1,i} x_{2,i} & \sum x_{2,i}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum x_{1,i} Y_i \\ \sum x_{2,i} Y_i \end{bmatrix}$$

$$\begin{bmatrix} 4 & 3.1781 & 10 \\ 3.1781 & 3.6092 & 10.2273 \\ 10 & 10.2273 & 30 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 7.0213 \\ 6.2636 \\ 19.0280 \end{bmatrix}$$

$$[a_0\ a_1\ a_2]' = [7.0213\ 6.2636\ 19.0280]'$$

$a = e^{a_0} = 1.2332$, $b = a_1 = -1.4259$, $c = a_2 = 1.0505$, and

$$y = ax^b e^{cx} = 1.2332 \cdot x^{-1.4259} \cdot e^{1.0505x}.$$



Linearization of nonlinear relationships

Documents has been collected from

https://en.wikipedia.org/wiki/Curve_fitting

https://www.ece.mcmaster.ca/~xwu/part5.pdf