Uncertainty of Parabola Fitting Through the Origin by Least Square Method

Xingju Dang, Jun Sun, Yao Zhang, Wenliang Wu
Institute of Physics and Electronic Information Engineering, Zhaotong University, China

ABSTRACT: This paper proposes the concept of correlation coefficient of parabolic fitting through origin and presents the computational formulas of the quadratic coefficient and monomial coefficient of parabolic fitting through origin by the least square method, coupled with the computational formula of the corresponding A-class uncertainty. Finally, case study is carried out.

1 GENERAL INSTRUCTIONS

In experimental data processing, sometimes it is necessary to fit curves by the data, and parabola fitting through the origin is a simple way. There are few researches on the evaluation of the uncertainty of experimental data processing with the parabola. Based on the previous literature, analyze the A-class uncertainty of parabola fitting through the origin by the least square method.

Respectively obtain two sets of measured series of two physical quantities \( x \) and \( y \): 
\[
X = (x_1, x_2, \ldots, x_n) \quad \text{and} \quad Y = (y_1, y_2, \ldots, y_n).
\]
Assume that \( y \) is the quadratic function \( y = ax^2 + bx \), use the least square method to calculate \( a \) and \( b \) based on measured series \( X \) and \( Y \), shown as follows:

Calculate the residual sum of squares
\[
Q = \sum_{i=1}^{n} \left( y_i - ax_i^2 - bx_i \right)^2. \tag{1}
\]

Calculate the partial derivatives of \( Q \) with respect to \( a \) and \( b \):
\[
\frac{\partial Q}{\partial a} = \sum_{i=1}^{n} 2 \left( y_i - ax_i^2 - bx_i \right) (-2ax_i) = \sum_{i=1}^{n} \left( 2ax_i^2 + bx_i - y_i \right),
\]
\[
\frac{\partial Q}{\partial b} = \sum_{i=1}^{n} 2 \left( y_i - ax_i^2 - bx_i \right) (-x_i) = \sum_{i=1}^{n} \left( a\sum_{i=1}^{n} x_i^2 + bx_i^2 - y_i \right).
\]

Make them equal to 0 and obtain the binary linear equation set:
\[
\begin{cases}
a \sum_{i=1}^{n} x_i^4 + b \sum_{i=1}^{n} x_i^3 = \sum_{i=1}^{n} y_i x_i^2, \\
a \sum_{i=1}^{n} x_i^3 + b \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} y_i x_i.
\end{cases}
\]

Obtain
\[
\begin{cases}
a = \frac{D_1}{D}, \\
b = \frac{D_2}{D},
\end{cases}
\tag{2}
\]

\[
D = \begin{vmatrix}
\sum_{i=1}^{n} x_i^4 & \sum_{i=1}^{n} x_i^3 \\
\sum_{i=1}^{n} x_i^3 & \sum_{i=1}^{n} x_i^2
\end{vmatrix},
\]
\[
D_1 = \begin{vmatrix}
\sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^3 \\
\sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i^2
\end{vmatrix},
D_2 = \begin{vmatrix}
\sum_{i=1}^{n} x_i y_i & \sum_{i=1}^{n} x_i^2 \\
\sum_{i=1}^{n} x_i & \sum_{i=1}^{n} x_i y_i
\end{vmatrix}. \tag{3}
\]

Where

Discuss the uncertainty of \( a \) and \( b \).

2 THE RESIDUAL SUM OF SQUARES AND CORRELATION COEFFICIENTS OF PARABOLA FITTING THROUGH THE ORIGIN

Put equation (2) into (1) and obtain
\[
Q = \sum_{i=1}^{n} \left( y_i - ax_i^2 - bx_i \right)^2 = \sum_{i=1}^{n} y_i^2 + a^2 \sum_{i=1}^{n} x_i^4 + b^2 \sum_{i=1}^{n} x_i^2 - 2ay_i x_i^2 - 2bx_i y_i + 2abx_i^2 = \sum_{i=1}^{n} y_i^2 + \frac{1}{D} \left( D_1 \sum_{i=1}^{n} x_i^4 + D_2 \sum_{i=1}^{n} x_i^2 + \frac{D_1}{D} \right) + \frac{2}{D} \left( D_1 D_2 \sum_{i=1}^{n} x_i^3 - D_1 \sum_{i=1}^{n} x_i y_i - D_2 \sum_{i=1}^{n} x_i y_i \right).
\]

The following equation is obtained from equation

(3)
\[ D, D, \sum x_i^4 - DD, \sum x_i^2 y_i = \]
\[ D_1 \left( \sum x_i^4 \sum x_i^2 y_i \sum x_i^4 \sum x_i y_i \sum x_i^2 y_i \right) = \]
\[ D_2 \frac{\sum x_i^4}{\sum x_i^2} + 0 \sum x_i y_i \sum x_i^4 - \sum x_i^2 y_i \sum x_i^4 = -D_1 \sum x_i^4, \]

So
\[ Q = \sum y_i^2 + \frac{1}{D}(-D_1 \sum x_i^2 + D_2 \sum x_i^2 - 2DD_2 \sum x_i y_i). \]

Since
\[ -D_1 \sum x_i^2 + D_2 \sum x_i^2 - 2DD_2 \sum x_i y_i = \]
\[ D_2 (D_2 \sum x_i^2 - D_2 \sum x_i y_i) - D_1 \sum x_i^2 - DD_2 \sum x_i y_i = \]
\[ D_2 \left( \sum x_i^4 \sum x_i^2 y_i \sum x_i^2 - \sum x_i^2 y_i \sum x_i^4 \right) - \]
\[ D_2 \sum x_i^4 - DD_2 \sum x_i y_i = \]
\[ -D, D_2 \sum x_i^4 - D_2 \sum x_i y_i = \]
\[ -D, D_2 \sum x_i^4 - D_2 \sum x_i^2 y_i = \]
\[ \sum x_i^4 \sum x_i^2 y_i \sum x_i^2 - \sum x_i^2 y_i \sum x_i^4 \]

Thus
\[ Q = \sum y_i^2 - (a \sum x_i^2 y_i + b \sum x_i y_i) = \]
\[ \sum y_i^2 \left( 1 - \frac{a \sum x_i^2 y_i + b \sum x_i y_i}{\sum y_i} \right). \]

The equation of the residual sum of squares of parabola through the origin fit by the least square method is shown as follows.

Here prove \( D \geq 0 \):
\[ D = \sum_{i=1}^{n} x_i^4 - \left( \sum_{i=1}^{n} x_i^2 \right)^2 = \]
\[ \sum_{i=1}^{n} x_i^4 - \sum_{i=1}^{n} \sum_{j=1}^{n} x_i x_j = \sum_{i=1}^{n} \sum_{j=1}^{n} \left( x_i^4 x_j^2 - x_i x_j^3 \right), \]
\[ x_i^4 x_j^2 - x_i x_j^3 = 0; \]

When \( i=j \), when \( i \neq j \),
\[ x_i^4 x_j^2 + x_j^4 x_i^2 - 2x_i x_j^3 = (x_i^2 x_j - x_j x_i^2)^2 \geq 0. \]

\[ 0 \leq a \sum x_i^2 y_i + b \sum x_i y_i \leq \sum y_i^2 : \]
\[ a \sum x_i^2 y_i + b \sum x_i y_i = \frac{1}{D} (D_2 \sum x_i^2 y_i + D_2 \sum x_i y_i) = \]
\[ \frac{1}{D} \left( \sum x_i^2 y_i \sum x_i^2 + \sum x_i^2 \sum x_i y_i \sum x_i y_i \right) = \]
\[ \frac{1}{D} \left( \sum x_i^4 \sum x_i y_i - \sum x_i^2 y_i \sum x_i y_i \sum x_i y_i \right) = \]
\[ \sum x_i^2 y_i \sum x_i^2 + \sum x_i^2 \sum x_i y_i - 2 \sum x_i^2 y_i \sum x_i y_i \sum x_i y_i. \]

So \( D \geq 0 \).

Prove when \( D \neq 0 \),
\[ \sum x_i^2 y_i \sum x_i^2 + \sum x_i^2 \sum x_i y_i \sum x_i y_i \geq \]
\[ 2 \sqrt{\sum x_i^2 y_i \sum x_i^2 \sum x_i y_i} \sum x_i y_i \sum x_i y_i \sum x_i y_i \geq \]
\[ 2 \sum x_i^2 y_i \sum x_i y_i \sum x_i y_i \sum x_i y_i \sum x_i y_i \geq 2 \sum x_i^2 y_i \sum x_i y_i \sum x_i y_i. \]

As \( D > 0 \) in the proof procedure,
\[ 0 \leq a \sum x_i^2 y_i + b \sum x_i y_i. \]

And due to the non-negativity of \( Q \) and \( \sum y_i^2 \),
\[ a \sum x_i^2 y_i + b \sum x_i y_i \leq \sum y_i^2. \]

So when \( D \neq 0 \), make
\[ r = \frac{a \sum x_i^2 y_i + b \sum x_i y_i}{\sum y_i^2} \]

It is called the correlation coefficient of parabola fitting through the origin. It is a number whose absolute value is less than or equal to 1 and the dimension is 1. The process of parabola fitting through the origin has some properties similar to the linear correlation coefficient. Thus, the residual sum of squares can be simplified as

\[ Q = \sum y_i^2 (1 - r^2) \]
3 A-CLASS STANDARD UNCERTAINTY OF a

Ignore the A-class uncertainty of each \( x_i \). Assume that the A-class standard uncertainty of each \( y_i \) is the same and mutually independent,

\[
u_A(y_i) = \sqrt{Q/(n-2)} = \sqrt{\sum \left(1 - r^2\right)/\left(n-2\right)}.
\]

Where \( n-2 \) means the degree of freedom.

As

\[
\frac{\partial}{\partial y_i} D_1 = \frac{\partial}{\partial y_i} \left( \sum_{j=1}^{n} y_j x_i^2 \sum_{j=1}^{n} x_j - \sum_{j=1}^{n} y_j x_i \sum_{j=1}^{n} x_j^3 \right) = x_i^2 \sum_{j=1}^{n} x_j^2 - x_i \sum_{j=1}^{n} x_j^3,
\]

So

\[
u_A(a) = \frac{1}{D} \left( \sum_{j=1}^{n} (\frac{\partial a}{\partial y_i} \nu_A(y_i))^2 \right) = \frac{1}{D} \left( \sum_{j=1}^{n} y_j^2 \left( \sum_{j=1}^{n} x_j^2 - x_i \sum_{j=1}^{n} x_j^3 \right) \right)^2 = \frac{1}{(n-2)D} \sum y_j^2 \sum x_j^2.
\]

Namely, A-class standard uncertainty of \( a \).

4 A-CLASS STANDARD UNCERTAINTY OF b

As

\[
\frac{\partial}{\partial y_i} D_2 = x_i \sum_{j=1}^{n} x_j - x_i \sum_{j=1}^{n} x_j^3
\]

So

\[
u_A(b) = \frac{1}{D} \left( \sum_{j=1}^{n} (\frac{\partial b}{\partial y_i} \nu_A(y_i))^2 \right) = \frac{1}{D} \left( \sum_{j=1}^{n} y_j \left( \sum_{j=1}^{n} x_j^3 - x_i \sum_{j=1}^{n} x_j^3 \right) \right)^2 = \frac{1}{(n-2)D} \sum y_j^2 \sum x_j^2.
\]

Namely, A-class standard uncertainty of \( b \).

5 APPLICATION

Illustrate the application of the equations by the following two examples.

5.1 Rotation with uniform acceleration

Set the position of photoelectric door A as 0.0300 m and that of photoelectric door B. The distance between A and B is measured as follows:

\[s = (0.7900, 0.8200, 0.8500, 0.8800, 0.9100, 0.9400, 0.9700, 1.0000, 1.0300, 1.0600, 1.0900, 1.1200) \text{m} \]

Successively measure the measured series of the corresponding freefall time of the steel ball (take the average by multiple measurements):

\[t = (0.369750, 0.377430, 0.384612, 0.392043, 0.399143, 0.405953, 0.412960, 0.419955, 0.426473, 0.433120, 0.439850, 0.446285) \text{s} \]

Assume that the following relationship exists between \( s \) and \( t \):

\[s = v_0 t + \frac{1}{2} g t^2 \]

Comparing \( y = s/t \), use the least square method to conduct linear fitting for \( y \) and \( t \), and ignore the uncertainty of class B to obtain the local gravitational acceleration \( g = (9.779 \pm 0.030) \text{m/s}^2 \), the speed of small ball through door A \( v_0 = (0.3286 \pm 0.0061) \text{m/s} \), and the correlation coefficient of \( y \) and \( t \) is \( r = 0.999955 \).

Carry out calculation by this method to obtain \( D = 0.00228776 \, s^6 \), \( D_1 = 0.0111808 \, \text{m/s}^4 \), \( D_2 = 0.00754037 \, \text{m/s}^3 \), \( a = 4.88724 \, \text{m/s}^2 \), \( b = 0.32960 \, \text{m/s} \), \( r = 0.999999 \), \( u_A(a) = 0.0149 \, \text{m/s} \), \( u_A(b) = 0.0165 \, \text{m/s} \). Therefore, \( g = (9.774 \pm 0.030) \, \text{m/s}^2 \), \( v_0 = (0.3296 \pm 0.0062) \, \text{m/s} \).

Compare the two methods and the best estimated values of the measured results are consistent within experimental error. The uncertainties of class A are similar.

5.2 Freefall

Make the JM-3 revolving inertia instrument produced by Xi’an University of Technology rotate with uniform acceleration under the effect of gravity moment of the weight and the frictional resisting moment of rotating shaft. Use the HMS-3 all-purpose computerized millisecond meter from Xi’an University of Technology to measure time and angular acceleration. The measured series is shown as follows:

\[\theta = (0, 1, 2, 3, 4, 5, 6, 7, 8, 9) \pi \text{ rad};\]

\[t = (0, 1.5370, 2.7183, 3.7334, 4.6429, 5.4836, 6.2664, 7.0089, 7.7159, 8.3956) \text{s};\]

\[\beta = (0.438, 0.411, 0.389, 0.371, 0.358, 0.344) \text{ rad/s}^2.\]
\( \beta \) is calculated by the instrument according to the equation

\[
\beta_i = \frac{2\pi}{t_{i+1}^3 t_{i+3}^2 - t_{i+3}^3 t_{i+1}^2} ((i+2)t_{i+1} - i t_{i+3}), \quad i = 1, 2, \ldots, 6
\]

To obtain the best estimated value of angular acceleration in the rotating process, the following three methods could be used.

Method 1: regard the 6 \( \beta_i \) calculated by the instrument as the direct measured values, and take the average as the best estimated value of \( \beta \). The experimental standard deviation of the mean value is regarded as the A-class uncertainty of \( \beta \). Obtain

\[
\beta = (0.3637 \pm 0.0070) \text{ rad/ s}^2.
\]

Method 2: set \( y = \frac{\theta}{t} \), conduct linear fitting for \( y \) and \( t \) and obtain \( \beta = (0.3837 \pm 0.0079) \text{ rad/ s}^2 \).

Method 3: use the method in this paper to conduct parabolic fitting for \( \theta \) and \( t \) and obtain:

\[
D = 209714.87 \text{ s}^6, D_1 = 38141.112 \text{ rad/s}^4,
D_2 = 388840.32 \text{ rad/s}^5, r = 0.999989,
\alpha = 0.18187 \text{ rad/s}^2, u_A(\alpha) = 0.0035 \text{ rad/s}^2
\]

And obtain

In method 2 and 3, the first set of data (\( \theta = 0, t = 0 \)) is not involved in fitting. The comparison of these three methods indicates that the best estimated values of method 1 and 2 are similar, but the uncertainty of method 2 is significantly lower than that of method 1, and that of method 3 is far different from method 1 and 2, and its uncertainty is slightly lower than that of method 2.

The above two examples indicate that this paper provides a feasible method for the fitting of the experimental data. The Class B uncertainty measured by this method and the evaluation of the combined uncertainty, coupled with the comparison with the linear fitting method, will be discussed in other paper.

REFERENCES