An Approach to . . .

Numerical Differentiation of Experimental Data

A procedure is described for the calculation of first and second derivatives from tabulated data. A method for estimating the error of the derivatives is presented.

The experimental measurement of a rate often requires that observed quantities be differentiated. Thus, in the measurement of drying rate, the total mass of a wet solid is usually observed as a function of time, the rate of loss in weight being calculated by differentiating the data. To determine the heat flux at a point in a two-fluid heat exchanger the mean temperature of one stream can be observed at several points along the heat transfer surface, the heat flux being calculated from a heat balance involving the first derivative of temperature with respect to position.

In each of these examples and in many similar ones the primary experimental measurements are subject to error and the differential quotients which are calculated from the data are therefore uncertain. The error in the calculated quantities should be no greater than that caused by the error in the primary observations. The method of calculation used should not introduce additional errors per se. The probable error of the derived quantity should be estimated from the scatter of the observed readings to find the reliability of the result.

Probably the most common method of finding the value of a first derivative consists of plotting the data on graph paper, drawing a line through the scattered points with the help of a French curve, and then drawing a tangent to this curve. The slope of this tangent is obviously affected by the mechanics of drawing the curved line and the tangent. Although mechanical devices involving mirrors or glass rods have been used to aid in locating the tangent, the fact remains that a curve through the points must be drawn, and it is by no means certain that the curve selected arbitrarily will be the most probable one.

Numerical methods have also been used to avoid the graphical constructions. One of the simplest of these is based on Newton’s interpolation formula (4). The values of y are listed along with the corresponding values of x, which must be equally spaced. The successive differences in y are then calculated, and from these the coefficients of x in a power series representing the derivative can be computed. The method involves the selection of a polynomial which passes exactly through each of the experimental points, followed by its differentiation.

If, for example, five points should be used the empirical equation relating y and x would be of the fourth degree, containing five disposable constants.

This mathematical accuracy is seldom warranted when the data are affected by experimental errors. Instead of using an empirical equation with as many constants as there are data points it is better to employ a simpler mathematical formula, choosing the fewer constants in it in such a way that the formula will represent the data as faithfully as possible. The theory of random errors leads to methods of selecting the best values of the arbitrary constants such that the sum of the squares of the deviations of the experimental measured quantity, y, from the value y’ calculated from the empirical formula shall be as small as is possible for a formula of the type chosen.

Sherwood and Reed (4) give a formula for the first derivative which is based on these least-squares methods and which they attribute to Douglass and Avakian. The formula gives the value of dy/dx at the center point of seven, the values of x being equally spaced by an amount Δx. The seven values of y, corresponding to the seven equally spaced values of x, are -3, -2, -1, 0, 1, 2, 3. The method assumes that a fourth-degree polynomial is sufficiently flexible to represent the functional relationship between y and x.

\[
\frac{dy}{dx} = \left( \frac{397}{1512} \Sigma y - \frac{7}{216} \Sigma xy \right) / \Delta x \quad (1)
\]

While the formula can be applied almost as easily as the data can be plotted and a curve drawn through them, calculation of the difference of the two sums sometimes requires that a calculating machine be employed. Furthermore, no equations were given from which the probable error of dy/dx could be calculated. It therefore appeared desirable to develop simpler methods based on the general least squares procedure in the hope that the resulting expressions could be used more rapidly, though with some sacrifice in accuracy, and that simple formulas for the standard deviation could be obtained.

Better estimates of the derivative and its error can always be obtained by the use of a larger number of points and with the knowledge, if available, of the scatter in the original data. The scheme presented here owes its merit to the fact that the derivative at the central point of a group of five points can be computed quickly and easily and on estimate of the error can be readily obtained.

Use of Least-Squares Equations

Although many different functions may be used to obtain a least-squares approximation to experimental data, it has been found that the use of Gram orthogonal polynomials (2) is especially convenient. For this type of approximation, y’, the computed value of the dependent variable, is given by

\[
y' = A_0P_0(x) + A_1P_1(x) + \ldots \quad (2)
\]

\[X = (x - x_c)\Delta x \] and \[x_c\] represents the central entry in the table of y and x. The value of y’ is to be distinguished from y, which is the corresponding measured
value. The first few polynomial functions are

\[ P_0(X) = 1 \]
\[ P_1(X) = \frac{1}{2} X \]
\[ P_2(X) = \frac{1}{2} (X^2 - 2) \]
\[ P_3(X) = \frac{1}{4} (5X^3 - 17X) \]

The higher-order functions can be found elsewhere (2).

When Gram polynomials are used with five evenly spaced points the coefficients are easily calculated by the use of Table I and Equation 3.

\[ A_r = \frac{1}{7r} \sum_{X=-r}^{X=r} y(X) P_r(X) \quad (3) \]

<p>| Table I. Values of Gram Polynomials in Equation 3 |
|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>X</th>
<th>P_0</th>
<th>P_1</th>
<th>P_2</th>
<th>P_3</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>1</td>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>y - 2</td>
</tr>
<tr>
<td>-1</td>
<td>1</td>
<td>-1</td>
<td>2</td>
<td>-2</td>
<td>2 - y - 1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>y_0</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-2</td>
<td>2y + 1</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>y + 2</td>
</tr>
<tr>
<td>( \gamma_r )</td>
<td>5</td>
<td>2</td>
<td>10</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The polynomials and coefficients, \( \gamma_r \), can easily be developed for larger odd numbers of evenly spaced points. Once the order of the curve has been decided upon the derivatives can be found by differentiating Equation 2. For a parabolic fit, these equations are listed in Table II.

<p>| Table II. Least-Squares Formulas for Derivatives |
|---|---|---|---|---|---|</p>
<table>
<thead>
<tr>
<th>X</th>
<th>( dy/dx )</th>
<th>( (dy/dx)^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>( \frac{-2XY_y}{10Ax} - \frac{2X^2Y_y}{7Ax} + \frac{42Y_y}{7Ax} )</td>
<td></td>
</tr>
<tr>
<td>-1</td>
<td>( \frac{-2XY_y}{10Ax} - \frac{2X^2Y_y}{7Ax} + \frac{22Y_y}{7Ax} )</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>( \frac{-2XY_y}{10Ax} - \frac{2X^2Y_y}{7Ax} )</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>( \frac{2XY_y}{10Ax} + \frac{2X^2Y_y}{7Ax} )</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>( \frac{2XY_y}{10Ax} + \frac{2X^2Y_y}{7Ax} - \frac{42Y_y}{7Ax} )</td>
<td></td>
</tr>
</tbody>
</table>

Note particularly the simple form of the formula for the derivative at the central point. In contrast with the Douglass-Avakian formula, its use does not require that a calculating machine be employed for evaluating the sums, because it is not necessary, as it is in Equation 1, to subtract two large numbers. Using the equation for \( dy/dx \) at the middle of the table, the derivative at a point can be computed more quickly numerically (and statistically more correctly) than it can graphically.

**Example.** The following data (3) give the partial pressure, \( y \), of di-tert-butyl peroxide as a function of time, \( x \), during decomposition to acetone and ethane.

<table>
<thead>
<tr>
<th>y, Min.</th>
<th>X, Min.</th>
<th>X</th>
<th>Xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>150.4</td>
<td>6</td>
<td>-2</td>
<td>-300.8</td>
</tr>
<tr>
<td>141.7</td>
<td>9</td>
<td>-1</td>
<td>-141.7</td>
</tr>
<tr>
<td>133.8</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>126.4</td>
<td>15</td>
<td>1</td>
<td>126.4</td>
</tr>
<tr>
<td>119.1</td>
<td>18</td>
<td>2</td>
<td>238.2</td>
</tr>
</tbody>
</table>

The derivative at a pressure of 133.8 mm. is given by:

\[ dy/dx = -77.9/(3)(10) = -2.597 \text{ mm./min.} \]

The first-order rate constant is

\[ k \times 10^4 = 19.40 \text{ min.}^{-1} \]

The Douglass-Avakian formula may also be used for comparison.

<table>
<thead>
<tr>
<th>y, Min.</th>
<th>X, Min.</th>
<th>X</th>
<th>Xy</th>
</tr>
</thead>
<tbody>
<tr>
<td>159.3</td>
<td>3</td>
<td>-3</td>
<td>-477.9</td>
</tr>
<tr>
<td>141.4</td>
<td>6</td>
<td>-2</td>
<td>-300.8</td>
</tr>
<tr>
<td>141.7</td>
<td>9</td>
<td>-1</td>
<td>-141.7</td>
</tr>
<tr>
<td>133.8</td>
<td>12</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>126.4</td>
<td>15</td>
<td>1</td>
<td>126.4</td>
</tr>
<tr>
<td>119.1</td>
<td>18</td>
<td>2</td>
<td>238.2</td>
</tr>
<tr>
<td>112.6</td>
<td>21</td>
<td>3</td>
<td>337.8</td>
</tr>
</tbody>
</table>

\[ dy/dx = \frac{(397)(-218.0) - (49)(-1526.6)}{(1512)(3)} = -2.593 \text{ mm./min.} \]

The first-order rate constant is

\[ k \times 10^4 = 19.38 \text{ min.}^{-1} \]

**Standard Error of Derivatives**

Worthing and Geffner (6) show that the standard deviation of a derived quantity, \( U \), is related to the standard deviations of the independent quantities, \( M \) and \( N \), to which it is related. If

\[ U = f(M, N) \quad (4) \]

it follows from straightforward statistical rules that

\[ \sigma_U^2 = \frac{\sum_{i=1}^{k} (y_i - \bar{y})^2}{k - p} \quad (5) \]

Note that it is assumed there is no uncertainty in the selection of the functional relationship between \( U, M, \) and \( N \); the error in \( U \), it is assumed, is due only to experimental errors in \( M \) and \( N \). Here, the problem is to find the value of \( \sigma \) for \( dy/dx \), which results from the standard deviations corresponding to the several experimentally observed, independent values of \( y \). Thus, applying Equation 5 and assuming that \( \sigma_y \) is the same for all five values of \( y \), for the point \( x \),

\[ \sigma_y = \sqrt{\sum_{i=1}^{5} \frac{(X_i - 10Ax)^2}{7Ax} + \frac{(X_i^2 - 7Ax)^2}{7Ax}} \quad (6) \]

Equation 7 can be reduced to the form

\[ \frac{dy/dx}{\Delta x} = \frac{\beta y}{\Delta x} \quad (7) \]

where \( \beta = 0.32 \) for \( X = 0 \), \( \beta = 0.62 \) for \( X = \pm 1 \) and \( \beta = 0.96 \) for \( x = \pm 2 \). For the second derivative

\[ \frac{dy/dx}{\Delta x} = \frac{0.54y}{(\Delta x)^3} \quad (8) \]

The problems are now how shall \( \sigma_y \) be determined and what does it represent. One of the assumptions involved in fitting a least-squares curve to some experimental data is that the function chosen to fit the data is in fact the true functional relationship between \( y \) and \( x \). It is selected perhaps on physical grounds, but does not fit the data exactly because the observed quantities are in error.

Notice that this is the same assumption upon which the derivation of Equation 5 is based. The term, \( \sigma_y \), is then the standard deviation of the experimental data. Assume for the present that this piece of information is not available from the experimental data as only a single value of \( y \) was measured at each value of \( x \). The standard deviation may then be calculated by Equation 9.

\[ \sigma_y^2 = \frac{\sum_{i=1}^{k} (y_i - \bar{y})^2}{k - p} \quad (9) \]
relationship is unknown, and the standard deviation calculated by Equation 9 is made up of the error in the original data and the error caused by the choice of an improper function to correlate the data. If repeated measurements of \( y \) were made at each value of \( x \) and the least-squares curve was fitted to the average \( y \) at each \( x \) an independent estimate of the standard deviation is available.

The assumption that the function used was the true functional relationship can be tested by comparing the standard deviation calculated by Equation 9 to that calculated from the repeated measurements (7). If the standard deviation calculated by Equation 9 is significantly larger, the function chosen was not the proper one. In this presentation the type of function has been chosen as a second-order polynomial because of the very simple equations obtained for this function. To retain this simplicity, the functional relationship will not be altered if it is found that the two standard deviations are significantly different. In many cases the standard deviation calculated by Equation 9 can be reduced by the transformation of the dependent variable \( y \) to some new variable such as \( \ln y \) or \( \exp y \). The data may be fitted using any number of transformations and the standard deviation calculated for each one. The best transformation to be used in fitting the data will naturally be the one having the smallest standard deviation.

In the following example (continued from the previous one) an estimate of the error of the derivative at the 99% confidence level is made.

**Example.** Find the estimated error at a pressure of 113.8 mm.

**Solution.**

From Equations 2 and 3 and Table I:

\[
y' = 134.3 - 7.8X + 0.25(X^2 - 2)
\]

<table>
<thead>
<tr>
<th>( X )</th>
<th>( y' )</th>
<th>( (y' - \bar{y})^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>-2</td>
<td>150.4</td>
<td>0.0</td>
</tr>
<tr>
<td>-1</td>
<td>141.9</td>
<td>0.04</td>
</tr>
<tr>
<td>0</td>
<td>133.8</td>
<td>0.0</td>
</tr>
<tr>
<td>1</td>
<td>126.3</td>
<td>0.01</td>
</tr>
<tr>
<td>2</td>
<td>119.2</td>
<td>0.00</td>
</tr>
</tbody>
</table>

The standard deviation is

\[ s_y = \left( \frac{0.06}{5 - 3} \right)^{1/2} \]

The standard deviation for \( dy/dx \) is now found from Equation 7 to be

\[ \frac{dy/dx}{dx} = \frac{(0.32)(0.03)^2}{(3)} = 0.018 \]

Using Student's \( t \)-distribution the estimated error at a 99% confidence level is found to be

\[ (9.92)(0.018) = 0.179 \]

The calculated value of the derivative may therefore be written as

\[ dy/dx = -2.597 \pm 0.179 \text{ mm./min.} \]

Because the error in the pressure measurement is small compared to the error in the derivative the rate constant may be written as

\[ k = (1.940 \pm 0.134) \times 10^{-1} \text{ min}^{-1} \]

**Suitability of Parabolic Approximation**

The question whether a second-degree polynomial is adequate for representing data may very well be raised, especially when the experimental methods are precise and, consequently, the resulting values of \( y \) and \( x \) fall nicely along a smooth curve. If this curve cannot be fitted perfectly by means of a parabola \( y' \) will have a value different from zero when the parabola is assumed, even though there may be no experimental error whatever in \( y \). The expected value of the derivative may not be so good as the data justify, owing to the fact that the slope of the best parabolic approximation may be different from the slope of a justifiable but more complicated equation fitting the data.

In some cases the original variables may be changed to new ones giving a more nearly linear function relationship, or the numerical procedure may be applied to the residual quantity left after subtracting a statistically definite approximating function from the observed values of \( y \). If the residuals are to be used, however, no advantage is gained unless the approximation function contains powers of \( x \) higher than the second, for the numerical procedures represented in Table II are capable of taking care of any parabolic functional relationship between the original variables.

**Example.** Calculate the value of \( dy/dx \) in the previous example using a change of variable to make the original data correspond more closely to linear variations.

**Solution.**

Letting \( z = \ln y \), \( dy/dx \) will be calculated numerically and the desired derivative will be found as follows:

\[ dy/dx = d(\exp z)/dx = yX^{2}f \]

<table>
<thead>
<tr>
<th>( y )</th>
<th>( z = \ln y )</th>
<th>( X )</th>
<th>( X^{2}f )</th>
</tr>
</thead>
<tbody>
<tr>
<td>150.4</td>
<td>5.01330</td>
<td>-2</td>
<td>-10.02660</td>
</tr>
<tr>
<td>141.7</td>
<td>4.90571</td>
<td>-1</td>
<td>-4.95371</td>
</tr>
<tr>
<td>133.8</td>
<td>4.80604</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>126.4</td>
<td>4.83944</td>
<td>1</td>
<td>4.83944</td>
</tr>
<tr>
<td>119.2</td>
<td>4.73996</td>
<td>2</td>
<td>9.55992</td>
</tr>
</tbody>
</table>

\[ y = \exp (zX^{2}f) \]

The desired derivative is

\[ dy/dx = (-0.58095)(133.8) = -2.591 \text{ mm./min.} \]

The reaction rate constant is \( k \times 10^{3} = 19.36 \text{ min}^{-1} \) Now the standard deviation is found using Equation 3 and Table I.

\[ x' = 4.89656 \pm 0.05809X + 0.00005(X^{2} - 2) \]

The standard deviation is

\[ s_{dy/dx} = \left( \frac{0.32)(0.32)(133.8)}{(3)} \right) = 0.015 \]

For this case the estimated error at a 99% confidence level is found to be

\[ \text{Error} = (9.92)(0.015) = 0.149 \]

and the derivative may be written as

\[ dy/dx = -2.591 \pm 0.149 \text{ mm./min.} \]

The rate constant becomes

\[ k = (1.936 \pm 0.111) \times 10^{-4} \text{ mm./min.} \]

The fact that the standard deviation is 5.7% of the estimated derivative according to the second computation compared with 6.9% in the first indicates that the larger value was due partly to the unsuitability of a parabola for representing the data.

**Acknowledgment**

The authors appreciate the helpful comments of F. H. Tingey.

**Literature Cited**


Received for review April 17, 1958
Accepted August 24, 1959
relationship is unknown, and the standard deviation calculated by Equation 9 is made up of the error in the original data and the error caused by the choice of an improper function to correlate the data. If repeated measurements of y were made at each value of x and the least-squares curve was fitted to the average y at each x an independent estimate of the standard deviation is available.

The assumption that the function used was the true functional relationship can be tested by comparing the standard deviation calculated by Equation 9 to that calculated from the repeated measurements (7). If the standard deviation calculated by Equation 9 is significantly larger, the function chosen was not the proper one. In this presentation the type of function has been chosen as a second-order polynomial because of the very simple equations obtained for this function. To retain this simplicity, the functional relationship will not be altered if it is found that the two standard deviations are significantly different. In many cases the standard deviation calculated by Equation 9 can be reduced by the transformation of the dependent variable y to some new variable such as ln y, y^{1/2}, or exp y. The data may be fitted using any number of transformations and the standard deviation calculated for each one. The best transformation to be used in fitting the data will naturally be the one having the smallest standard deviation.

In the following example (continued from the previous one) an estimate of the standard deviation calculated by Equation 9 will have a value different from zero when the parabola is assumed, even though there may be no experimental error whatever in y. The expected value of the derivative may not be so good as the data justify, owing to the fact that the slope of the best parabolic approximation may be different from the slope of a justifiable but more complicated equation fitting the data.

In some cases the original variables may be changed to new ones giving a more nearly linear function relationship, or the numerical procedure may be applied to the residual quantity left after subtracting a statistically definite approximating function from the observed values of y. If the residuals are to be used, however, no advantage is gained unless the approximation function contains powers of x higher than the second, for the numerical procedures represented in Table II are capable of taking care of any parabolic functional relationship between the original variables.

**Example.** Calculate the value of dy/dx in the previous example using a change of variable to make the original data correspond more closely to linear variations.

**Solution.**

The substitution of ln y for y in the previous table will produce an improvement. Letting z = ln y, dy/dx will be calculated numerically and the desired derivative will be found as follows:

\[
\frac{dy}{dx} = \frac{d(exp z)}{dx} = y \frac{dz}{dx}
\]

Using the data in Table I and Equation 2 the dy/dx is found from Equation 7 to be

\[
\frac{dy}{dx} = \frac{(0.32)(0.03)}{(3)} = 0.018
\]

The standard deviation for dy/dx is now found from Equation 9 to be

\[
\frac{S_{dy/dx}}{dy/dx} = \frac{(0.32)(0.03)}{(3)} = 0.018
\]

The standard deviation for the estimated error at a 99% confidence level is found to be

\[
S_y = \sqrt{\frac{t^2}{n-3}} = \sqrt{\frac{12.7}{3}} = 2.591
\]

The reaction rate constant is

\[
\frac{dy}{dx} = \frac{(-0.58095)(133.8)}{(3)} = -2.591 \text{ mm./min.}
\]

The rate constant becomes

\[
k = (1.940 \pm 0.134) \times 10^{-2} \text{ min.}^{-1}
\]

**Suitability of Parabolic Approximation**

The question whether a second-degree polynomial is adequate for representing data may very well be raised, especially when the experimental methods are precise and, consequently, the resulting values of y and x fall nicely along a smooth curve. If this curve cannot be fitted perfectly by means of a parabola dy/dx will have a value different from zero when the parabola is assumed, even though there may be no experimental error whatever in y. The expected value of the derivative may not be so good as the data justify, owing to the fact that the slope of the best parabolic approximation may be different from the slope of a justifiable but more complicated equation fitting the data.

**Example.** Using Table I and Equation 2 the dy/dx is found to be

\[
\frac{dy}{dx} = \frac{(9.92)(0.015)}{(3)} = 0.0015
\]

The desired derivative is

\[
\frac{dy}{dx} = \frac{(-0.58095)(133.8)}{(3)} = -2.591 \text{ mm./min.}
\]

The rate constant becomes

\[
k = (1.936 \pm 0.111) \times 10^{-3} \text{ min.}^{-1}
\]

The fact that the standard deviation is 5.7% of the estimated derivative according to the second computation compared with 6.9% in the first indicates that the larger value was due partly to the unsuitability of a parabola for representing the data.

**Acknowledgment**

The authors appreciate the helpful comments of F. H. Tingey.

**Literature Cited**


For seven pts

\[ p_0 = 1 \]

\[ p_1 = x/3 \]

\[ p_2 = (x^2 - 4)/5 \]

\[ p_3 = (x^3 - 7x)/6 \]

\[ \gamma_0 = 7 \]

\[ \gamma_1 = 28/9 \]

\[ \gamma_2 = 84/25 \]

\[ \gamma_3 = 6 \]

\[ \frac{d^3 y}{dx^3} = \frac{A_3}{\Delta x^3} \]

\[ A_3 = \frac{1}{\gamma_3} \sum_{x=-3}^{x=+3} y(x) p_3(x) \]