

Computational Mathematics/Information Technology

Worksheet 6

Cubic splines and Least squares

Cubic Splines.

Problem 1 The linear spline is a poor approximation to $f(x)$ away from the knots (*data points*). We now consider the construction of a cubic spline through the following data points.

x_k	0	1	2	3	4
$y_k = f(x_k)$	6.00000	4.62091	1.75156	0.03002	1.03907

For this set of data the k th component of the spline is given by:

$$s_k(x) = a_k + b_k(x - x_k) + c_k(x - x_k)^2 + d_k(x - x_k)^3 \quad \text{for } k = 0, 1, \dots, 3.$$

- Each component has to pass through the points at the ends of the interval on which it is defined. Therefore,

$$\begin{aligned} s_0(x_0) &= f(x_0), \\ s_0(x_1) &= f(x_1), \\ s_1(x_1) &= f(x_1), \\ s_1(x_2) &= f(x_2), \\ s_2(x_2) &= f(x_2), \\ s_2(x_3) &= f(x_3), \\ s_3(x_3) &= f(x_3), \\ s_3(x_4) &= f(x_4). \end{aligned}$$

- The derivative of each component has to match the derivative of the next component at their common knot. Therefore,

$$\begin{aligned} s'_0(x_1) &= s'_1(x_1), \\ s'_1(x_2) &= s'_2(x_2), \\ s'_2(x_3) &= s'_3(x_3). \end{aligned}$$

- The second derivative of each component has to match the second derivative of the next component at their common knot. Therefore,

$$\begin{aligned} s''_0(x_1) &= s''_1(x_1), \\ s''_1(x_2) &= s''_2(x_2), \\ s''_2(x_3) &= s''_3(x_3). \end{aligned}$$

You can see that we have 14 conditions here. However, we have 16 coefficient values (a_k, b_k, c_k and d_k for $k = 0, 1, 2$ and 3) so we need two more conditions.

- The **natural** cubic spline is defined by specifying two extra conditions, namely,

$$\begin{aligned} s_0(x_0)'' &= 0 \\ s_3(x_4)'' &= 0 \end{aligned}$$

Finding the coefficients

The calculation of the coefficients using the above conditions can be formalized by generalizing the approach used in the notes. You are not expected at the moment to be able to derive or know the general method however you may be able to spot from the following example how the method works when you have more data points. (*It may help you to understand things if you try deriving some of the expressions written here*)

- a_0, a_1, a_2, a_3 and c_0

Using the above data and the conditions $s_0(x_0) = f(x_0)$, $s_1(x_1) = f(x_1)$, $s_2(x_2) = f(x_2)$ and $s_3(x_3) = f(x_3)$, write down the coefficients a_k :

$$\begin{aligned} a_0 &= \dots\dots\dots, & a_1 &= \dots\dots\dots, \\ a_2 &= \dots\dots\dots, & a_3 &= \dots\dots\dots \end{aligned}$$

Using $s_0''(x_0) = 0$ show that $c_0 = 0$.

- c_1, c_2 and c_3

It **can** be shown (*for the case of equally spaced x values, as in this example*) that the remaining conditions imply that the coefficients c_1, c_2 and c_3 satisfy the matrix equation

$$\begin{pmatrix} 4 & 1 & 0 \\ 1 & 4 & 1 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix} = \frac{3}{h} \begin{pmatrix} \delta_1 - \delta_0 \\ \delta_2 - \delta_1 \\ \delta_3 - \delta_2 \end{pmatrix} \quad (1)$$

where, δ_k is given by:

$$\delta_k = \frac{f(x_{k+1}) - f(x_k)}{h}$$

and h is the interval of x spacings, which in this example is given by $h = 1$.

Thus

$$\delta_k = f(x_{k+1}) - f(x_k)$$

By direct hand calculation complete the following:

$$\frac{3}{h} \begin{pmatrix} \delta_1 - \delta_0 \\ \delta_2 - \delta_1 \\ \delta_3 - \delta_2 \end{pmatrix} = \frac{3}{h} \begin{pmatrix} \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \end{pmatrix} = \begin{pmatrix} \dots\dots\dots \\ \dots\dots\dots \\ \dots\dots\dots \end{pmatrix}$$

You now have the right-hand side of Eq(1), hence use Excel to solve this matrix equation for c_1, c_2 and c_3 .

$$c_1 = \dots\dots\dots, \quad c_2 = \dots\dots\dots$$

$$\text{and } c_3 = \dots\dots\dots.$$

- d_0, d_1, d_2 and d_3

The d s in terms of the c s can be shown to be:

$$d_0 = \frac{c_1}{3} = \dots\dots\dots d_1 = \frac{c_2 - c_1}{3} = \dots\dots\dots$$

$$d_2 = \frac{c_3 - c_2}{3} = \dots\dots\dots d_3 = -\frac{c_3}{3} = \dots\dots\dots$$

- b_0, b_1, b_2 and b_3

The b s are given by the following (*in our example $h = 1$*)

$$b_0 = \delta_0 - c_0h - d_0h^2 = \delta_0 - c_0 - d_0 = \dots\dots\dots$$

$$b_1 = \delta_1 - c_1h - d_1h^2 = \delta_1 - c_1 - d_1 = \dots\dots\dots$$

$$b_2 = \delta_2 - c_2h - d_2h^2 = \delta_2 - c_2 - d_2 = \dots\dots\dots$$

$$b_3 = \delta_3 - c_3h - d_3h^2 = \delta_3 - c_3 - d_3 = \dots\dots\dots$$

You now have all the coefficients. Hence write down the cubic spline:

$$S(x) = \begin{cases} s_0(x) = \dots\dots\dots \\ s_1(x) = \dots\dots\dots \\ s_2(x) = \dots\dots\dots \\ s_3(x) = \dots\dots\dots \end{cases}$$

Plot the data points and the cubic spline

This follows the procedure from the previous worksheet.

- Construct two columns containing just the data points.
- Plot a graph of the data points using the chart type that consists only of points (a scatter graph)
- Write a VBA function module for $S(x)$. (See Worksheet 5 for hints)
- Construct a column of x values on your worksheet in the range $x = 0$ to $x = 1$ in steps of 0.02. (Use auto fill facility to produce these values)
- Use your VBA function to construct a corresponding column of $S(x)$ values next to the column of x values. Note: one of these will be needed for the question sheet.
- Copy and Paste these points to the graph using Paste Special.
- Change the plot of these to a line plot using the method of the previous worksheet.

Least Squares Fitting

Single variable linear fitting

Problem 2

Find the least squares linear fit $y = m_1x + c$, to the given data points as follows:

$x =$	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$y =$	1	0.997	0.987	0.970	0.946	0.915	0.877	0.831	0.777	0.714	0.642

- Enter the x values in cells A1:A11 and the y values in cells B1:B11.
- Highlight cells A13:B13
- enter **=LINEST(B1:B11,A1:A11)** followed by **Ctrl-Shift-Enter**

Cell A13 should now contain m_1 and cell B13 contain c .

The linear fit is given by: $y =$

Problem 3 (plot data and fit)

- Set up a column of y values using $y = m_1x + c$:
 - Enter **=\$A\$13*A1+\$B\$13** in cell C1
 - Copy this down to cell C11.
 - Cells C1 to C11 should now contain the predicted y values.
- Plot the original data:
 - Highlight cells A1 to B11.
 - Plot the data points.
- Plot the straight line fit on the same graph:
 - Highlight and Copy the y values C1 to C11
 - Click on the chart and Past the values just copied. The points will be plotted with the same x values as the last set of data. (*this is the simplest way of adding data to a chart and is appropriate when the new data has exactly the same set of x values as the existing set.*)
 - To change the points to a plot joining the points with straight lines.

Does this look like a good approximation?

Polynomial Fitting

Problem 4 (Add polynomial Trendline)

- To add a polynomial fit to the data on the chart Excel provides us with the Add Trendline option.
 - Activate the chart by left clicking just inside the boundary. “Chart Tools” should now be visible at the top of Excel.
 - Make sure the original data points on the graph are selected.
 - Below “Chart tools” should be the options Design, Layout, and Format. select Layout.
 - Click on the “Trendline” menu and select “More Trendline Options” from the bottom.
 - Select Polynomial with Order 3.

The “best fit” cubic should now be shown on your chart. At this stage the equation of the cubic is not known. To calculate it proceed as follows:

Problem 5 (*Cubic fit*)

To fit the polynomial $y = p_3(x) = m_3x^3 + m_2x^2 + m_1x + c$ to the data in problem 2:

- Construct three columns, one for each of x , x^2 and x^3 and one column for the y data:
 - Copy the x data from A1:A11 to A17:A27
 - Enter =A17^2 in cell B17. Copy down to B27
 - Enter =A17^3 in cell C17. Copy down to C27
 - Copy the y data from B1:B11 to D17:D27
- Find the parameters using LINEST:
 - Highlight cells A30:D30
 - enter =**LINEST(D17:D27,A17:C27)** followed by **Ctrl-Shift-Enter**

Cell A30 should now contain m_3 , cell B30 m_2 , cell C30 m_1 and cell D30 contain c .

The cubic polynomial fit is given by: $y = p_3(x) =$

Excel will do this for you. On the “Trendline options” menu was a box to check to give this polynomial on the graph. Use this to check your answer.

Problem 6 (*Plot data and cubic*)

Create a new plot. On it plot the original data points, and a curve of the cubic that you calculated by evaluating the cubic with a step length of 0.02 for x .

Does this look like a good approximation?

Problem 7 (*coefficient of determination*)

LINEST also provides us with statistics that enable us to judge how good an approximation we are making. The easiest indicator to consider is the coefficient of determination, denoted r^2 , where r is the correlation coefficient. Recall the closer r^2 is to +1 the better the fit. To make LINEST display the statistics we need to use:

=**LINEST**(*known y*, *known x*, **TRUE**, **TRUE**)

- To find r^2 for Problem 2. (*linear fit*)
 - LINEST requires five rows and two columns to output all its results. Highlight E1:F5.
 - Enter =**LINEST(B1:B11,A1:A11, TRUE, TRUE)** followed by: **Ctrl-Shift-Enter**
 - The first row contains m_1 and c as before. The second row contains the standard errors. The third row contains r^2 and “SEy”.

$r^2 =$

- To find r^2 for Problem 5. (*cubic fit*)
 - Since we are determining four parameters, m_1 , m_2 , m_3 and c , LINEST will require five rows and four columns to output all its results. Highlight H1:K5
 - Recall that the y -values are in D17:D27 and the x -values are in A17:C27, thus: Enter =**LINEST(D17:D27,A17:C27, TRUE, TRUE)** followed by: **Ctrl-Shift-Enter**

– As before r^2 is in the third row, first column:

$$r^2 = \dots\dots\dots$$

Are the above values of r^2 in agreement with your comments regarding the goodness of fit based on the two graphs above?

Problem 8 (*an application*)

The above data is in fact generated from the function $f(x) = \frac{x}{\tan x}$ with $f(0) = 1$ and rounded to three decimal places. We can use the polynomial in many application, for example to approximate the function at an intermediate point or as here, to evaluate an integral.

Evaluate

$$\int_0^1 p_3(x) dx \quad \text{to approximate} \quad \int_0^1 \frac{x}{\tan x} dx$$

by entering $p_3(x)$ into Derive and carrying out the definite integral.

$$\int_0^1 p_3(x) dx = \dots\dots\dots$$

Enter the function $\frac{x}{\tan x}$ into Derive and hence carry out an approximate integration, obtaining the answer to 4 decimal places.

$$\int_0^1 f(x)(x) dx \approx \dots\dots\dots$$

Assessment Exercises Worksheet 6

Splines and Least Squares Fitting

Surname:.....

Forename:.....

Group No¹.

--

For the Actuarial Science students this sheet is to be handed in at the associated lab session. For Maths students it should be handed in at the second of the two associated lab sessions. Late submission is only allowed with the permission of the lab tutor.

As appropriate express answers correct to four decimal places

1. For Problem 1 write down the cubic spine found in the worksheet:

$$S(x) = \begin{cases} \dots\dots\dots & 0 \leq x \leq 1 \\ \dots\dots\dots & 1 \leq x \leq 2 \\ \dots\dots\dots & 2 \leq x \leq 3 \\ \dots\dots\dots & 3 \leq x \leq 4 \end{cases}$$

Find $S(2.5) = \dots\dots\dots$

2. For Problem 2 write down the linear fit: $y = \dots\dots\dots$

3. From Problem 5 write down the cubic fit

$$y = p_3(x) = \dots\dots\dots$$

4. From Problem 7 write down the values of r^2 for:

(a) the linear fit in Problem 2: $\dots\dots\dots$

(b) the cubic fit in Problem 5: $\dots\dots\dots$

Which of these, (a) or (b), gives the best fit to the given data: $\dots\dots\dots$

¹Actuarial Science Students only