NAG Fortran Library Routine Document

E02BAF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of ***bold italicised*** terms and other implementation-dependent details.

1 Purpose

E02BAF computes a weighted least-squares approximation to an arbitrary set of data points by a cubic spline with knots prescribed by the user. Cubic spline interpolation can also be carried out.

2 Specification

```
      SUBROUTINE E02BAF(M, NCAP7, X, Y, W, LAMDA, WORK1, WORK2, C, SS, IFAIL)
      INTEGER           M, NCAP7, IFAIL
      real              X(M), Y(M), W(M), LAMDA(NCAP7), WORK1(M),
     1                  WORK2(4*NCAP7), C(NCAP7), SS
```

3 Description

This routine determines a least-squares cubic spline approximation $s(x)$ to the set of data points (x_r, y_r) with weights w_r, for $r = 1, 2, \ldots, m$. The value of NCAP7 $= \bar{n} + 7$, where $\bar{n}$ is the number of intervals of the spline (one greater than the number of interior knots), and the values of the knots $\lambda_5, \lambda_6, \ldots, \lambda_{\bar{n}+3}$, interior to the data interval, are prescribed by the user.

$s(x)$ has the property that it minimizes θ, the sum of squares of the weighted residuals ϵ_r, for $r = 1, 2, \ldots, m$, where

$$\epsilon_r = w_r(y_r - s(x_r)).$$

The routine produces this minimizing value of θ and the coefficients $c_1, c_2, \ldots, c_q$, where $q = \bar{n} + 3$, in the B-spline representation

$$s(x) = \sum_{i=1}^{q} c_i N_i(x).$$

Here $N_i(x)$ denotes the normalised B-spline of degree 3 defined upon the knots $\lambda_i, \lambda_{i+1}, \ldots, \lambda_{i+4}$.

In order to define the full set of B-splines required, eight additional knots $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ and $\lambda_{\bar{n}+4}, \lambda_{\bar{n}+5}, \lambda_{\bar{n}+6}, \lambda_{\bar{n}+7}$ are inserted automatically by the routine. The first four of these are set equal to the smallest x_r and the last four to the largest x_r.

The representation of $s(x)$ in terms of B-splines is the most compact form possible in that only $\bar{n} + 3$ coefficients, in addition to the $\bar{n} + 7$ knots, fully define $s(x)$.

The method employed involves forming and then computing the least-squares solution of a set of m linear equations in the coefficients c_i $(i = 1, 2, \ldots, \bar{n} + 3)$. The equations are formed using a recurrence relation for B-splines that is unconditionally stable (Cox (1972a), de Boor (1972)), even for multiple (coincident) knots. The least-squares solution is also obtained in a stable manner by using orthogonal transformations, viz. a variant of Givens rotations (Gentleman (1974) and Gentleman (1973)). This requires only one equation to be stored at a time. Full advantage is taken of the structure of the equations, there being at most four non-zero values of $N_i(x)$ for any value of x and hence at most four coefficients in each equation.

For further details of the algorithm and its use see Cox (1974), Cox (1975b) and Cox and Hayes (1973).

Subsequent evaluation of $s(x)$ from its B-spline representation may be carried out using E02BBF. If derivatives of $s(x)$ are also required, E02BCF may be used. E02BDF can be used to compute the definite integral of $s(x)$.

4 References

Cox M G (1972a) The numerical evaluation of B-splines *J. Inst. Math. Appl.* **10** 134–149

Cox M G (1974) A data-fitting package for the non-specialist user *Software for Numerical Mathematics* (ed D J Evans) Academic Press

Cox M G (1975b) Numerical methods for the interpolation and approximation of data by spline functions *PhD Thesis* City University, London

Cox M G and Hayes J G (1973) Curve fitting: a guide and suite of algorithms for the non-specialist user *NPL Report NAC26* National Physical Laboratory

de Boor C (1972) On calculating with B-splines *J. Approx. Theory* **6** 50–62

Gentleman W M (1974) Algorithm AS 75. Basic procedures for large sparse or weighted linear least-squares problems *Appl. Statist.* **23** 448–454

Gentleman W M (1973) Least-squares computations by Givens transformations without square roots *J. Inst. Math. Applic.* **12** 329–336

Schoenberg I J and Whitney A (1953) On Polya frequency functions III *Trans. Amer. Math. Soc.* **74** 246–259

5 Parameters

1: M – INTEGER *Input*

On entry: the number m of data points.

Constraint: M $\geq$ MDIST ≥ 4, where MDIST is the number of distinct x values in the data.

2: NCAP7 – INTEGER *Input*

On entry: $\bar{n}+7$, where $\bar{n}$ is the number of intervals of the spline (which is one greater than the number of interior knots, i.e., the knots strictly within the range x_1 to x_m) over which the spline is defined.

Constraint: $8 \leq$ NCAP7 $\leq$ MDIST $+ 4$, where MDIST is the number of distinct x values in the data.

3: X(M) – ***real*** array *Input*

On entry: the values x_r of the independent variable (abscissa), for $r = 1, 2, \ldots, m$.

Constraint: $x_1 \leq x_2 \leq \cdots \leq x_m$.

4: Y(M) – ***real*** array *Input*

On entry: the values y_r of the dependent variable (ordinate), for $r = 1, 2, \ldots, m$.

5: W(M) – ***real*** array *Input*

On entry: the values w_r of the weights, for $r = 1, 2, \ldots, m$. For advice on the choice of weights, see the E02 Chapter Introduction.

Constraint: $\mathrm{W}(r) > 0$, for $r = 1, 2, \ldots, m$.

6: LAMDA(NCAP7) – ***real*** array *Input/Output*

On entry: LAMDA(i) must be set to the $(i-4)$th (interior) knot, λ_i, for $i = 5, 6, \ldots, \bar{n}+3$.

Constraint: X(1) < LAMDA(5) $\leq$ LAMDA(6) $\leq \cdots \leq$ LAMDA(NCAP7 – 4) < X(M).

On exit: the input values are unchanged, and LAMDA(i), for $i = 1, 2, 3, 4$, NCAP7 – 3, NCAP7 – 2, NCAP7 – 1, NCAP7 contains the additional (exterior) knots introduced by the routine. For advice on the choice of knots, see Section 3.3 of the E02 Chapter Introduction.

7: WORK1(M) – ***real*** array *Workspace*

8: WORK2(4*NCAP7) – ***real*** array *Workspace*

9: C(NCAP7) – ***real*** array *Output*

On exit: the coefficient c_i of the B-spline $N_i(x)$, for $i = 1, 2, \ldots, \bar{n} + 3$. The remaining elements of the array are not used.

10: SS – ***real*** *Output*

On exit: the residual sum of squares, θ.

11: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

On exit: IFAIL $= 0$ unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value -1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value -1 or 1 is used it is essential to test the value of IFAIL on exit.**

6 Error Indicators and Warnings

If on entry IFAIL $= 0$ or -1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

IFAIL $= 1$

The knots fail to satisfy the condition

X(1) $<$ LAMDA(5) $\leq$ LAMDA(6) $\leq \cdots \leq$ LAMDA(NCAP7 – 4) $<$ X(M).

Thus the knots are not in correct order or are not interior to the data interval.

IFAIL $= 2$

The weights are not all strictly positive.

IFAIL $= 3$

The values of X(r), for $r = 1, 2, \ldots,$ M are not in non-decreasing order.

IFAIL $= 4$

NCAP7 < 8 (so the number of interior knots is negative) or NCAP7 $>$ MDIST $+ 4$, where MDIST is the number of distinct x values in the data (so there cannot be a unique solution).

IFAIL $= 5$

The conditions specified by Schoenberg and Whitney (1953) fail to hold for at least one subset of the distinct data abscissae. That is, there is no subset of NCAP7 – 4 strictly increasing values, X(R(1)), X(R(2)), $\ldots$, X(R(NCAP7 – 4)), among the abscissae such that

$$X(R(1)) < \text{LAMDA}(1) < X(R(5)),$$
$$X(R(2)) < \text{LAMDA}(2) < X(R(6)),$$
$$\vdots$$
$$X(R(\text{NCAP7}-8)) < \text{LAMDA}(\text{NCAP7}-8) < X(R(\text{NCAP7}-4)).$$

This means that there is no unique solution: there are regions containing too many knots compared with the number of data points.

7 Accuracy

The rounding errors committed are such that the computed coefficients are exact for a slightly perturbed set of ordinates $y_r + \delta y_r$. The ratio of the root-mean-square value for the δy_r to the root-mean-square value of the y_r can be expected to be less than a small multiple of $\kappa \times m \times$***machine precision***, where κ is a condition number for the problem. Values of κ for 20–30 practical data sets all proved to lie between 4.5 and 7.8 (see Cox (1975b)). (Note that for these data sets, replacing the coincident end knots at the end-points x_1 and x_m used in the routine by various choices of non-coincident exterior knots gave values of κ between 16 and 180. Again see Cox (1975b) for further details.) In general we would not expect κ to be large unless the choice of knots results in near-violation of the Schoenberg–Whitney conditions.

A cubic spline which adequately fits the data and is free from spurious oscillations is more likely to be obtained if the knots are chosen to be grouped more closely in regions where the function (underlying the data) or its derivatives change more rapidly than elsewhere.

8 Further Comments

The time taken by the routine is approximately $\mathrm{C} \times (2m + \bar{n} + 7)$ seconds, where C is a machine-dependent constant.

Multiple knots are permitted as long as their multiplicity does not exceed 4, i.e., the complete set of knots must satisfy $\lambda_i < \lambda_{i+4}$, for $i = 1, 2, \ldots, \bar{n} + 3$, (see Section 6). At a knot of multiplicity one (the usual case), $s(x)$ and its first two derivatives are continuous. At a knot of multiplicity two, $s(x)$ and its first derivative are continuous. At a knot of multiplicity three, $s(x)$ is continuous, and at a knot of multiplicity four, $s(x)$ is generally discontinuous.

The routine can be used efficiently for cubic spline interpolation, i.e., if $m = \bar{n} + 3$. The abscissae must then of course satisfy $x_1 < x_2 < \cdots < x_m$. Recommended values for the knots in this case are $\lambda_i = x_{i-2}$, for $i = 5, 6, \ldots, \bar{n} + 3$.

9 Example

Determine a weighted least-squares cubic spline approximation with five intervals (four interior knots) to a set of 14 given data points. Tabulate the data and the corresponding values of the approximating spline, together with the residual errors, and also the values of the approximating spline at points half-way between each pair of adjacent data points.

The example program is written in a general form that will enable a cubic spline approximation with $\bar{n}$ intervals ($\bar{n} - 1$ interior knots) to be obtained to m data points, with arbitrary positive weights, and the approximation to be tabulated. Note that E02BBF is used to evaluate the approximating spline. The program is self-starting in that any number of data sets can be supplied.

9.1 Program Text

Note: the listing of the example program presented below uses ***bold italicised*** terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*     E02BAF Example Program Text
*     Mark 15 Revised.  NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          MMAX, NC7MAX
      PARAMETER        (MMAX=200,NC7MAX=50)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real             FIT, SS, XARG
      INTEGER          IFAIL, IWGHT, J, M, NCAP, R
*     .. Local Arrays ..
      real             C(NC7MAX), LAMDA(NC7MAX), W(MMAX), WORK1(MMAX),
     +                 WORK2(4*NC7MAX), X(MMAX), Y(MMAX)
*     .. External Subroutines ..
      EXTERNAL         E02BAF, E02BBF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'E02BAF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
   20 READ (NIN,*,END=100) M
      IF (M.GT.0 .AND. M.LE.MMAX) THEN
         READ (NIN,*) NCAP, IWGHT
         IF (NCAP+7.LE.NC7MAX) THEN
            IF (NCAP.GT.1) READ (NIN,*) (LAMDA(J),J=5,NCAP+3)
            DO 40 R = 1, M
               IF (IWGHT.EQ.1) THEN
                  READ (NIN,*) X(R), Y(R)
                  W(R) = 1.0e0
               ELSE
                  READ (NIN,*) X(R), Y(R), W(R)
               END IF
   40       CONTINUE
            IFAIL = 0
*
            CALL E02BAF(M,NCAP+7,X,Y,W,LAMDA,WORK1,WORK2,C,SS,IFAIL)
*
            WRITE (NOUT,*)
            WRITE (NOUT,*)
     +        '  J       LAMDA(J+2)     B-spline coeff C(J)'
            WRITE (NOUT,*)
            J = 1
            WRITE (NOUT,99998) J, C(1)
            DO 60 J = 2, NCAP + 2
               WRITE (NOUT,99999) J, LAMDA(J+2), C(J)
   60       CONTINUE
            WRITE (NOUT,99998) NCAP + 3, C(NCAP+3)
            WRITE (NOUT,*)
            WRITE (NOUT,99997) 'Residual sum of squares = ', SS
            WRITE (NOUT,*)
            WRITE (NOUT,*) 'Cubic spline approximation and residuals'
            WRITE (NOUT,*)
            WRITE (NOUT,*)
     +       '  R   Abscissa     Weight   Ordinate     Spline  Residual'
            WRITE (NOUT,*)
            DO 80 R = 1, M
               IFAIL = 0
*
               CALL E02BBF(NCAP+7,LAMDA,C,X(R),FIT,IFAIL)
*
               WRITE (NOUT,99995) R, X(R), W(R), Y(R), FIT, FIT - Y(R)
               IF (R.LT.M) THEN
                  XARG = 0.5e0*(X(R)+X(R+1))
*
                  CALL E02BBF(NCAP+7,LAMDA,C,XARG,FIT,IFAIL)
*
```

```
                  WRITE (NOUT,99996) XARG, FIT
               END IF
   80       CONTINUE
            GO TO 20
         END IF
      END IF
  100 STOP
*
99999 FORMAT (1X,I3,F15.4,F20.4)
99998 FORMAT (1X,I3,F35.4)
99997 FORMAT (1X,A,e12.2)
99996 FORMAT (1X,F14.4,F33.4)
99995 FORMAT (1X,I3,4F11.4,e10.2)
      END
```

9.2 Program Data

```
E02BAF Example Program Data
  14
   5   2
     1.50
     2.60
     4.00
     8.00
     0.20      0.00      0.20
     0.47      2.00      0.20
     0.74      4.00      0.30
     1.09      6.00      0.70
     1.60      8.00      0.90
     1.90      8.62      1.00
     2.60      9.10      1.00
     3.10      8.90      1.00
     4.00      8.15      0.80
     5.15      7.00      0.50
     6.17      6.00      0.70
     8.00      4.54      1.00
    10.00      3.39      1.00
    12.00      2.56      1.00
```

9.3 Program Results

```
 E02BAF Example Program Results

   J       LAMDA(J+2)     B-spline coeff C(J)

   1                            -0.0465
   2         0.2000              3.6150
   3         1.5000              8.5724
   4         2.6000              9.4261
   5         4.0000              7.2716
   6         8.0000              4.1207
   7        12.0000              3.0822
   8                             2.5597

 Residual sum of squares =     0.18E-02

 Cubic spline approximation and residuals

   R   Abscissa     Weight   Ordinate     Spline  Residual

   1     0.2000     0.2000     0.0000    -0.0465 -0.47E-01
         0.3350                           1.0622
   2     0.4700     0.2000     2.0000     2.1057  0.11E+00
         0.6050                           3.0817
   3     0.7400     0.3000     4.0000     3.9880 -0.12E-01
         0.9150                           5.0558
   4     1.0900     0.7000     6.0000     5.9983 -0.17E-02
         1.3450                           7.1376
   5     1.6000     0.9000     8.0000     7.9872 -0.13E-01
         1.7500                           8.3544
```

```
     6      1.9000      1.0000      8.6200      8.6348  0.15E-01
            2.2500                              9.0076
     7      2.6000      1.0000      9.1000      9.0896 -0.10E-01
            2.8500                              9.0353
     8      3.1000      1.0000      8.9000      8.9125  0.12E-01
            3.5500                              8.5660
     9      4.0000      0.8000      8.1500      8.1321 -0.18E-01
            4.5750                              7.5592
    10      5.1500      0.5000      7.0000      6.9925 -0.75E-02
            5.6600                              6.5010
    11      6.1700      0.7000      6.0000      6.0255  0.26E-01
            7.0850                              5.2292
    12      8.0000      1.0000      4.5400      4.5315 -0.85E-02
            9.0000                              3.9045
    13     10.0000      1.0000      3.3900      3.3928  0.28E-02
           11.0000                              2.9574
    14     12.0000      1.0000      2.5600      2.5597 -0.35E-03
```