NAG Library Routine Document

G02BNF

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

G02BNF computes Kendall and/or Spearman nonparametric rank correlation coefficients for a set of data; the data array is overwritten with the ranks of the observations.

2 Specification

```
 SUBROUTINE G02BNF(N, M, X, LDX, ITYPE, RR, LDRR, KWORKA, KWORKB, WORK1,
1                  WORK2, IFAIL)
 INTEGER           N, M, LDX, ITYPE, LDRR, KWORKA(N), KWORKB(N), IFAIL
 double precision  X(LDX,M), RR(LDRR,M), WORK1(M), WORK2(M)
```

3 Description

The input data consists of n observations for each of m variables, given as an array

$$[x_{ij}], \qquad i = 1, 2, \ldots, n(n \geq 2), j = 1, 2, \ldots, m(m \geq 2),$$

where x_{ij} is the ith observation of the jth variable.

The quantities calculated are:

(a) Ranks

For a given variable, j say, each of the n observations, $x_{1j}, x_{2j}, \ldots, x_{nj}$, has associated with it an additional number, the 'rank' of the observation, which indicates the magnitude of that observation relative to the magnitudes of the other $n-1$ observations on that same variable.

The smallest observation for variable j is assigned the rank 1, the second smallest observation for variable j the rank 2, the third smallest the rank 3, and so on until the largest observation for variable j is given the rank n.

If a number of cases all have the same value for the given variable, j, then they are each given an 'average' rank, e.g., if in attempting to assign the rank $h+1$, k observations were found to have the same value, then instead of giving them the ranks

$$h+1, h+2, \ldots, h+k,$$

all k observations would be assigned the rank

$$\frac{2h+k+1}{2}$$

and the next value in ascending order would be assigned the rank

$$h+k+1.$$

The process is repeated for each of the m variables.

Let y_{ij} be the rank assigned to the observation x_{ij} when the jth variable is being ranked. The actual observations x_{ij} are replaced by the ranks y_{ij}.

(b) Nonparametric rank correlation coefficients

(i) Kendall's tau:

$$R_{jk} = \frac{\sum_{h=1}^{n}\sum_{i=1}^{n} \operatorname{sign}(y_{hj} - y_{ij}) \operatorname{sign}(y_{hk} - y_{ik})}{\sqrt{[n(n-1) - T_j][n(n-1) - T_k]}}, \qquad j,k = 1,2,\ldots,m,$$

where $\operatorname{sign} u = 1$ if $u > 0$,

$\operatorname{sign} u = 0$ if $u = 0$,

$\operatorname{sign} u = -1$ if $u < 0$,

and $T_j = \sum t_j(t_j - 1)$, where t_j is the number of ties of a particular value of variable j, and the summation is over all tied values of variable j.

(ii) Spearman's:

$$R^*_{jk} = \frac{n(n^2-1) - 6\sum_{i=1}^{n}(y_{ij} - y_{ik})^2 - \frac{1}{2}(T^*_j + T^*_k)}{\sqrt{[n(n^2-1) - T^*_j][n(n^2-1) - T^*_k]}}, \qquad j,k = 1,2,\ldots,m,$$

where $T^*_j = \sum t_j(t_j^2 - 1)$, t_j being the number of ties of a particular value of variable j, and the summation being over all tied values of variable j.

4 References

Siegel S (1956) *Non-parametric Statistics for the Behavioral Sciences* McGraw–Hill

5 Parameters

1: N – INTEGER *Input*

On entry: n, the number of observations or cases.

Constraint: $N \geq 2$.

2: M – INTEGER *Input*

On entry: m, the number of variables.

Constraint: $M \geq 2$.

3: X(LDX,M) – ***double precision*** array *Input/Output*

On entry: $X(i,j)$ must be set to x_{ij}, the value of the ith observation on the jth variable, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

On exit: $X(i,j)$ contains the rank y_{ij} of the observation x_{ij}, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

4: LDX – INTEGER *Input*

On entry: the first dimension of the array X as declared in the (sub)program from which G02BNF is called.

Constraint: $LDX \geq N$.

5: ITYPE – INTEGER *Input*

On entry: the type of correlation coefficients which are to be calculated.

ITYPE $= -1$
: Only Kendall's tau coefficients are calculated.

ITYPE = 0
Both Kendall's tau and Spearman's coefficients are calculated.

ITYPE = 1
Only Spearman's coefficients are calculated.

Constraint: ITYPE = −1, 0 or 1.

6: RR(LDRR,M) – ***double precision*** array *Output*

On exit: the requested correlation coefficients.

If only Kendall's tau coefficients are requested (ITYPE = −1), RR(j,k) contains Kendall's tau for the jth and kth variables.

If only Spearman's coefficients are requested (ITYPE = 1), RR(j,k) contains Spearman's rank correlation coefficient for the jth and kth variables.

If both Kendall's tau and Spearman's coefficients are requested (ITYPE = 0), the upper triangle of RR contains the Spearman coefficients and the lower triangle the Kendall coefficients. That is, for the jth and kth variables, where j is less than k, RR(j,k) contains the Spearman rank correlation coefficient, and RR(k,j) contains Kendall's tau, for $j,k = 1,2,\ldots,m$.

(Diagonal terms, RR(j,j), are unity for all three values of ITYPE.)

7: LDRR – INTEGER *Input*

On entry: the first dimension of the array RR as declared in the (sub)program from which G02BNF is called.

Constraint: LDRR $\geq$ M.

8: KWORKA(N) – INTEGER array *Workspace*
9: KWORKB(N) – INTEGER array *Workspace*
10: WORK1(M) – ***double precision*** array *Workspace*
11: WORK2(M) – ***double precision*** array *Workspace*

12: IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, −1 or 1. If you are unfamiliar with this parameter you should refer to Section 3.3 in the Essential Introduction 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, if you are 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

On entry, N < 2.

IFAIL = 2

On entry, M < 2.

IFAIL = 3

On entry, LDX < N,
or LDRR < M.

IFAIL = 4

On entry, ITYPE < −1,
or ITYPE > 1.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by G02BNF depends on n and m.

9 Example

This example reads in a set of data consisting of nine observations on each of three variables. The program then calculates and prints the rank of each observation, and both Kendall's tau and Spearman's rank correlation coefficients for all three variables.

9.1 Program Text

```
*     G02BNF Example Program Text
*     Mark 14 Revised. NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          M, N, LDX, LDRR
      PARAMETER        (M=3,N=9,LDX=N,LDRR=M)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, ITYPE, J
*     .. Local Arrays ..
      DOUBLE PRECISION RR(LDRR,M), WORK1(M), WORK2(M), X(LDX,M)
      INTEGER          KWORKA(N), KWORKB(N)
*     .. External Subroutines ..
      EXTERNAL         G02BNF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02BNF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) ((X(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      WRITE (NOUT,99999) 'Number of variables (columns) =', M
      WRITE (NOUT,99999) 'Number of cases     (rows)    =', N
      WRITE (NOUT,*)
      WRITE (NOUT,*) 'Data matrix is:-'
      WRITE (NOUT,*)
      WRITE (NOUT,99998) (J,J=1,M)
      WRITE (NOUT,99997) (I,(X(I,J),J=1,M),I=1,N)
      WRITE (NOUT,*)
      IFAIL = 1
      ITYPE = 0
*
      CALL G02BNF(N,M,X,LDX,ITYPE,RR,LDRR,KWORKA,KWORKB,WORK1,WORK2,
     +            IFAIL)
*
      IF (IFAIL.LT.0) THEN
         WRITE (NOUT,*)
         WRITE (NOUT,99999) ' ** G02BNF returned with IFAIL = ', IFAIL
      ELSE
         IF (IFAIL.NE.0) THEN
            WRITE (NOUT,99999) 'Routine fails, IFAIL =', IFAIL
```

```
         ELSE
            WRITE (NOUT,*) 'Matrix of ranks:-'
            WRITE (NOUT,99998) (J,J=1,M)
            WRITE (NOUT,99997) (I,(X(I,J),J=1,M),I=1,N)
            WRITE (NOUT,*)
            WRITE (NOUT,*) 'Matrix of rank correlation coefficients:'
            WRITE (NOUT,*) 'Upper triangle -- Spearman''s'
            WRITE (NOUT,*) 'Lower triangle -- Kendall''s tau'
            WRITE (NOUT,*)
            WRITE (NOUT,99998) (I,I=1,M)
            WRITE (NOUT,99997) (I,(RR(I,J),J=1,M),I=1,M)
         END IF
      END IF
*
99999 FORMAT (1X,A,I5)
99998 FORMAT (1X,3I12)
99997 FORMAT (1X,I3,3F12.4)
      END
```

9.2 Program Data

```
G02BNF Example Program Data
1.70      1.00      0.50
2.80      4.00      3.00
0.60      6.00      2.50
1.80      9.00      6.00
0.99      4.00      2.50
1.40      2.00      5.50
1.80      9.00      7.50
2.50      7.00      0.00
0.99      5.00      3.00
```

9.3 Program Results

```
 G02BNF Example Program Results

 Number of variables (columns) =    3
 Number of cases     (rows)    =    9

 Data matrix is:-

            1           2           3
   1      1.7000      1.0000      0.5000
   2      2.8000      4.0000      3.0000
   3      0.6000      6.0000      2.5000
   4      1.8000      9.0000      6.0000
   5      0.9900      4.0000      2.5000
   6      1.4000      2.0000      5.5000
   7      1.8000      9.0000      7.5000
   8      2.5000      7.0000      0.0000
   9      0.9900      5.0000      3.0000

 Matrix of ranks:-
            1           2           3
   1      5.0000      1.0000      2.0000
   2      9.0000      3.5000      5.5000
   3      1.0000      6.0000      3.5000
   4      6.5000      8.5000      8.0000
   5      2.5000      3.5000      3.5000
   6      4.0000      2.0000      7.0000
   7      6.5000      8.5000      9.0000
   8      8.0000      7.0000      1.0000
   9      2.5000      5.0000      5.5000

 Matrix of rank correlation coefficients:
 Upper triangle -- Spearman's
 Lower triangle -- Kendall's tau
```

```
             1            2            3
   1    1.0000       0.2246       0.1186
   2    0.0294       1.0000       0.3814
   3    0.1176       0.2353       1.0000
```