NAG Fortran Library Routine Document

G02BYF

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

G02BYF computes a partial correlation/variance-covariance matrix from a correlation or variance-covariance matrix computed by G02BXF.

2 Specification

```
SUBROUTINE G02BYF(M, NY, NX, ISZ, R, LDR, P, LDP, WK, IFAIL)
INTEGER            M, NY, NX, ISZ(M), LDR, LDP, IFAIL
real               R(LDR,M), P(LDP,NY), WK(NY*NX+NX*(NX+1)/2)
```

3 Description

Partial correlation can be used to explore the association between pairs of random variables in the presence of other variables. For three variables, y_1, y_2 and x_3, the partial correlation coefficient between y_1 and y_2 given x_3 is computed as:

$$\frac{r_{12} - r_{13}r_{23}}{\sqrt{(1-r_{13}^2)(1-r_{23}^2)}},$$

where r_{ij} is the product-moment correlation coefficient between variables with subscripts i and j. The partial correlation coefficient is a measure of the linear association between y_1 and y_2 having eliminated the effect due to both y_1 and y_2 being linearly associated with x_3. That is, it is a measure of association between y_1 and y_2 conditional upon fixed values of x_3. Like the full correlation coefficients the partial correlation coefficient takes a value in the range $(-1, 1)$ with the value 0 indicating no association.

In general, let a set of variables be partitioned into two groups Y and X with n_y variables in Y and n_x variables in X and let the variance-covariance matrix of all $n_y + n_x$ variables be partitioned into,

$$\begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}.$$

The the variance-covariance of Y conditional on fixed values of the X variables is given by:

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{yx}\Sigma_{xx}^{-1}\Sigma_{xy}.$$

The partial correlation matrix is then computed by standardising $\Sigma_{y|x}$,

$$\mathrm{diag}(\Sigma_{y|x})^{-\frac{1}{2}}\Sigma_{y|x}\,\mathrm{diag}(\Sigma_{y|x})^{-\frac{1}{2}}.$$

To test the hypothesis that a partial correlation is zero under the assumption that the data has an approximately Normal distribution a test similar to the test for the full correlation coefficient can be used. If r is the computed partial correlation coefficient then the appropriate t statistic is

$$r\sqrt{\frac{n - n_x - 2}{1 - r^2}},$$

which has approximately a Student's t-distribution with $n - n_x - 2$ degrees of freedom, where n is the number of observations from which the full correlation coefficents were computed.

4 References

Krzanowski W J (1990) *Principles of Multivariate Analysis* Oxford University Press

Morrison D F (1967) *Multivariate Statistical Methods* McGraw-Hill

Osborn J F (1979) *Statistical Exercises in Medical Research* Blackwell

Snedecor G W and Cochran W G (1967) *Statistical Methods* Iowa State University Press

5 Parameters

1: M – INTEGER *Input*

On entry: the number of variables in the variance-covariance/correlation matrix given in R.

Constraint: M ≥ 3.

2: NY – INTEGER *Input*

On entry: the number of Y variables, n_y, for which partial correlation coefficents are to be computed.

Constraint: NY ≥ 2.

3: NX – INTEGER *Input*

On entry: the number of X variables, n_x, which are to be considered as fixed.

Constraints:

NX ≥ 1,
NY + NX $\leq$ M.

4: ISZ(M) – INTEGER array *Input*

On entry: indicates which variables belong to set X and Y.

If ISZ$(i) < 0$ then the ith variable is a Y variable, for $i = 1, 2, \dots,$ M.

If ISZ$(i) > 0$ then the ith variable is a X variable.

If ISZ$(i) = 0$ then the ith variable is not included in the computations.

Constraints:

exactly NY elements of ISZ must be < 0,
exactly NX elements of ISZ must be > 0.

5: R(LDR,M) – ***real*** array *Input*

On entry: the variance-covariance or correlation matrix for the M variables as given by G02BXF. Only the upper triangle need be given.

Note: the matrix must be a full rank variance-covariance or correlation matrix and so be positive-definite. This condition is not directly checked by the routine.

6: LDR – INTEGER *Input*

On entry: the first dimension of the array R as declared in the (sub)program from which G02BYF is called.

Constraint: LDR $\geq$ M.

7: P(LDP,NY) – ***real*** array *Output*

On exit: the strict upper triangle of P contains the strict upper triangular part of the n_y by n_y partial correlation matrix. The lower triangle contains the lower triangle of the n_y by n_y partial variance-covariance matrix if the matrix given in R is a variance-covariance matrix. If the matrix given in R is a correlation matrix then the variance-covariance matrix is for standardised variables.

8: LDP – INTEGER *Input*

On entry: the first dimension of the array P as declared in the (sub)program from which G02BYF is called.

Constraint: LDP $\geq$ NY.

9: WK(NY*NX+NX*(NX+1)/2) – ***real*** array *Workspace*

10: 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, because for this routine the values of the output parameters may be useful even if IFAIL $\neq 0$ on exit, the recommended value is -1. **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, M < 3,
or NY < 2,
or NX < 1,
or NY $+$ NX $>$ M,
or LDR $<$ M,
or LDP $<$ NY.

IFAIL $= 2$

On entry, there are not exactly NY elements of ISZ < 0,
or there are not exactly NX elements of ISZ > 0.

IFAIL $= 3$

On entry, the variance-covariance/correlation matrix of the X variables, Σ_{xx}, is not of full rank. Try removing some of the X variables by setting the appropriate element of ISZ $= 0$.

IFAIL $= 4$

Either a diagonal element of the partial variance-covariance matrix, $\Sigma_{y|x}$, is zero and/or a computed partial correlation coefficient is greater than one. Both indicate that the matrix input in R was not positive-definite.

7 Accuracy

G02BYF computes the partial variance-covariance matrix, $\Sigma_{y|x}$, by computing the Cholesky factorization of Σ_{xx}. If Σ_{xx} is not of full rank the computation will fail. For a statement on the accuracy of the Cholesky factorization see F07GDF (SPPTRF/DPPTRF).

8 Further Comments

Models that represent the linear associations given by partial correlations can be fitted using the multiple regression routine G02DAF.

9 Example

Data, given by Osborn (1979), on the number of deaths, smoke (mg/m^3) and sulphur dioxide (parts/million) during an intense period of fog is input. The correlations are computed using G02BXF and the partial correlation between deaths and smoke given sulphur dioxide is computed using G02BYF. Both correlation matrices are printed using the routine X04CAF.

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.

```
*     G02BYF Example Program Text
*     Mark 17 Release. NAG Copyright 1995.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          LDX, MMAX, LDV
      PARAMETER        (LDX=20,MMAX=10,LDV=MMAX)
*     .. Local Scalars ..
      INTEGER          IFAIL, J, K, M, N, NX, NY
*     .. Local Arrays ..
      real             R(LDV,MMAX), STD(MMAX), V(LDV,MMAX),
     +                 WK(MMAX*MMAX), WT(LDX), X(LDX,MMAX), XBAR(MMAX)
      INTEGER          ISZ(MMAX)
*     .. External Subroutines ..
      EXTERNAL         G02BXF, G02BYF, X04CAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02BYF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, M
      IF (M.LE.MMAX .AND. N.LE.LDX) THEN
         READ (NIN,*) ((X(J,K),K=1,M),J=1,N)
*
*        Calculate correlation matrix
*
         IFAIL = -1
*
         CALL G02BXF('U',N,M,X,LDX,WT,XBAR,STD,V,LDV,R,IFAIL)
*
         IF (IFAIL.EQ.0) THEN
*
*        Print the correlation matrix
*
            WRITE (NOUT,*)
            CALL X04CAF('Upper','Non-unit',M,M,R,LDV,
     +                  'Correlation matrix',IFAIL)
            READ (NIN,*) NY, NX
            READ (NIN,*) (ISZ(J),J=1,M)
*
*           Calculate partial correlation matrix
*
            IFAIL = 0
```

```
*
            CALL G02BYF(M,NY,NX,ISZ,V,LDV,R,LDV,WK,IFAIL)
*
*           Print partial correlation matrix
*
            WRITE (NOUT,*)
            CALL X04CAF('Upper','Unit',NY,NY,R,LDV,
     +                  'Partial Correlation matrix',IFAIL)
         END IF
      END IF
      STOP
      END
```

9.2 Program Data

```
G02BYF Example Program Data
15 3
112 0.30 0.09
140 0.49 0.16
143 0.61 0.22
120 0.49 0.14
196 2.64 0.75
294 3.45 0.86
513 4.46 1.34
518 4.46 1.34
430 1.22 0.47
274 1.22 0.47
255 0.32 0.22
236 0.29 0.23
256 0.50 0.26
222 0.32 0.16
213 0.32 0.16

 2  1
-1 -1  1
```

9.3 Program Results

```
 G02BYF Example Program Results

 Correlation matrix
          1       2       3
 1   1.0000  0.7560  0.8309
 2           1.0000  0.9876
 3                   1.0000

 Partial Correlation matrix
          1       2
 1   1.0000 -0.7381
 2           1.0000
```