G03CCF – NAG Fortran Library Routine Document

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

G03CCF computes factor score coefficients from the result of fitting a factor analysis model by maximum likelihood as performed by G03CAF.

2 Specification

```
      SUBROUTINE G03CCF(METHOD, ROTATE, NVAR, NFAC, FL, LDFL, PSI, E, R,
     1                  LDR, FS, LDFS, WK, IFAIL)
      INTEGER           NVAR, NFAC, LDFL, LDR, LDFS, IFAIL
      real              FL(LDFL,NFAC), PSI(NVAR), E(NVAR), R(LDR,*),
     1                  FS(LDFS,NFAC), WK(NVAR)
      CHARACTER*1       METHOD, ROTATE
```

3 Description

A factor analysis model aims to account for the covariances among p variables, observed on n individuals, in terms of a smaller number, k, of unobserved variables or factors. The values of the factors for an individual are known as factor scores. G03CAF fits the factor analysis model by maximum likelihood and returns the estimated factor loading matrix, Λ, and the diagonal matrix of variances of the unique components, Ψ. To obtain estimates of the factors a p by k matrix of factor score coefficients, Φ, is formed. The estimated vector of factor scores, $\hat{f}$, is then given by:

$$\hat{f} = x^T \Phi,$$

where x is the vector of observed variables for an individual.

There are two commonly used methods of obtaining factor score coefficients.

The regression method:

$$\Phi = \Psi^{-1}\Lambda(I + \Lambda^T\Psi^{-1}\Lambda)^{-1},$$

and Bartlett's method:

$$\Phi = \Psi^{-1}\Lambda(\Lambda^T\Psi^{-1}\Lambda)^{-1}.$$

See Lawley and Maxwell [1] for details of both methods. In the regression method as given above, it is assumed that the factors are not correlated and have unit variance; this is true for models fitted by G03CAF. Further, for models fitted by G03CAF,

$$\Lambda^T\Psi^{-1}\Lambda = \Theta - I,$$

where Θ is the diagonal matrix of eigenvalues of the matrix S^*, as described in G03CAF.

The factors may be orthogonally rotated using an orthogonal rotation matrix, R, as computed by G03BAF. The factor scores for the rotated matrix are then given by ΛR.

4 References

[1] Lawley D N and Maxwell A E (1971) *Factor Analysis as a Statistical Method* Butterworths (2nd Edition)

5 Parameters

1: METHOD — CHARACTER*1 *Input*

On entry: indicates which method is to be used to compute the factor score coefficients.

If METHOD = 'R', then the regression method is used.
If METHOD = 'B', then Bartlett's method is used.

Constraint: METHOD = 'B' or 'R'.

2: ROTATE — CHARACTER*1 *Input*

On entry: indicates whether a rotation is to be applied.

If ROTATE = 'R', then a rotation will be applied to the coefficients and the rotation matrix, R, must be given in R.
If ROTATE = 'U', then no rotation is applied.

Constraint: ROTATE = 'R' or 'U'.

3: NVAR — INTEGER *Input*

On entry: the number of observed variables in the factor analysis, p.

Constraint: NVAR $\geq$ NFAC.

4: NFAC — INTEGER *Input*

On entry: the number of factors in the factor analysis, k.

Constraint: NFAC ≥ 1.

5: FL(LDFL,NFAC) — ***real*** array *Input*

On entry: the matrix of unrotated factor loadings, Λ, as returned by G03CAF.

6: LDFL — INTEGER *Input*

On entry: the first dimension of the array FL as declared in the (sub)program from which G03CCF is called.

Constraint: LDFL $\geq$ NVAR.

7: PSI(NVAR) — ***real*** array *Input*

On entry: the diagonal elements of Ψ, as returned by G03CAF.

Constraint: PSI(i) > 0.0, for $i = 1, 2, \ldots, p$.

8: E(NVAR) — ***real*** array *Input*

On entry: the eigenvalues of the matrix S^*, as returned by G03CAF.

Constraint: E(i) > 1.0, for $i = 1, 2, \ldots, p$.

9: R(LDR,$*$) — ***real*** array *Input*

Note: the second dimension of the array R must be at least 1 if ROTATE = 'U' and at least NFAC if ROTATE = 'R'.

On entry:

if ROTATE = 'R', then R must contain the orthogonal rotation matrix, R, as returned by G03BAF.
if ROTATE = 'U', then R need not be set.

10: LDR — INTEGER *Input*

On entry: the first dimension of the array R as declared in the (sub)program from which G03CCF is called.

Constraint: if ROTATE = 'R', LDR $\geq$ NFAC.

11: FS(LDFS,NFAC) — ***real*** array *Output*

On exit: the matrix of factor score coefficients, Φ. FS(i, j) contains the factor score coefficient for the jth factor and the ith observed variable, for $i = 1, 2, \ldots, p$; $j = 1, 2, \ldots, k$.

12: LDFS — INTEGER *Input*

On entry: the first dimension of the array FS as declared in the (sub)program from which G03CCF is called.

Constraint: LDFS $\geq$ NVAR.

13: WK(NVAR) — ***real*** array *Workspace*

14: IFAIL — INTEGER *Input/Output*

On entry: IFAIL must be set to 0, -1 or 1. For users not familiar with this parameter (described in Chapter P01) the recommended value is 0.

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

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 detected by the routine:

IFAIL = 1

On entry, NFAC $<$ 1,
or NVAR $<$ NFAC,
or LDFL $<$ NVAR,
or LDFS $<$ NVAR,
or ROTATE = 'R' and LDR $<$ NFAC,
or METHOD $\neq$ 'R' or 'B',
or ROTATE $\neq$ 'R' or 'U'.

IFAIL = 2

On entry, a value of PSI $\leq$ 0.0,
or a value of E $\leq$ 1.0.

7 Accuracy

Accuracy will depend on the accuracy requested when computing the estimated factor loadings using G03CAF.

8 Further Comments

To compute the factor scores using the factor score coefficients the values for the observed variables first need to be standardized by subtracting the sample means and, if the factor analysis is based upon a correlation matrix, dividing by the sample standard deviations. This may be performed using G03ZAF. The standardized variables are then post-multiplied by the factor score coefficients. This may be performed using routines from the F06 Chapter Introduction, for example F06YAF.

If principal component analysis is required the routine G03AAF computes the principal component scores directly. Hence, the factor score coefficients are not needed.

9 Example

The example is taken from Lawley and Maxwell [1]. The correlation matrix for 220 observations on six school subjects is input and a factor analysis model with two factors fitted using G03CAF. The factor score coefficients are computed using the regression method.

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.

```
*     G03CCF Example Program Text
*     Mark 15 Release. NAG Copyright 1991.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
      INTEGER          NMAX, MMAX, LWK
      PARAMETER        (NMAX=20,MMAX=10,LWK=400)
*     .. Local Scalars ..
      INTEGER          I, IFAIL, J, M, N, NFAC, NVAR
      CHARACTER        MATRIX, METHOD, WEIGHT
*     .. Local Arrays ..
      real             COM(MMAX), E(MMAX), FL(MMAX,MMAX), FS(MMAX,MMAX),
     +                 PSI(MMAX), R(MMAX,MMAX), STAT(4), WK(LWK),
     +                 WT(NMAX), X(NMAX,MMAX)
      INTEGER          IOP(5), ISX(MMAX), IWK(4*MMAX+2)
*     .. External Subroutines ..
      EXTERNAL         G03CAF, G03CCF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G03CCF Example Program Results'
*     Skip headings in data file
      READ (NIN,*)
      READ (NIN,*) MATRIX, WEIGHT, N, M, NVAR, NFAC
      IF (M.LE.MMAX .AND. (MATRIX.EQ.'C' .OR. MATRIX.EQ.'c' .OR. N.LE.
     +    NMAX)) THEN
         IF (MATRIX.EQ.'C' .OR. MATRIX.EQ.'c') THEN
            DO 20 I = 1, M
               READ (NIN,*) (X(I,J),J=1,M)
   20       CONTINUE
         ELSE
            IF (WEIGHT.EQ.'W' .OR. WEIGHT.EQ.'w') THEN
               DO 40 I = 1, N
                  READ (NIN,*) (X(I,J),J=1,M), WT(I)
   40          CONTINUE
            ELSE
               DO 60 I = 1, N
                  READ (NIN,*) (X(I,J),J=1,M)
   60          CONTINUE
            END IF
         END IF
         READ (NIN,*) (ISX(J),J=1,M)
         READ (NIN,*) (IOP(J),J=1,5)
         IFAIL = 1
*
         CALL G03CAF(MATRIX,WEIGHT,N,M,X,NMAX,NVAR,ISX,NFAC,WT,E,STAT,
     +               COM,PSI,R,FL,MMAX,IOP,IWK,WK,LWK,IFAIL)
*
         IF (IFAIL.EQ.0 .OR. IFAIL.GT.4) THEN
```

```
            WRITE (NOUT,*)
            WRITE (NOUT,*) ' Loadings, Communalities and PSI'
            WRITE (NOUT,*)
            DO 80 I = 1, NVAR
               WRITE (NOUT,99999) (FL(I,J),J=1,NFAC), COM(I), PSI(I)
   80       CONTINUE
            READ (NIN,*) METHOD
            IFAIL = 0
*
            CALL G03CCF(METHOD,'U',NVAR,NFAC,FL,MMAX,PSI,E,R,MMAX,FS,
     +                  MMAX,WK,IFAIL)
*
            WRITE (NOUT,*)
            WRITE (NOUT,*) ' Factor score coefficients'
            WRITE (NOUT,*)
            DO 100 I = 1, NVAR
               WRITE (NOUT,99999) (FS(I,J),J=1,NFAC)
  100       CONTINUE
         END IF
      END IF
      STOP
*
99999 FORMAT (2X,4F8.3)
      END
```

9.2 Program Data

```
G03CCF Example Program Data
'C' 'U' 220 6  6 2
1.000 0.439 0.410 0.288 0.329 0.248
0.439 1.000 0.351 0.354 0.320 0.329
0.410 0.351 1.000 0.164 0.190 0.181
0.288 0.354 0.164 1.000 0.595 0.470
0.329 0.320 0.190 0.595 1.000 0.464
0.248 0.329 0.181 0.470 0.464 1.000
  1    1    1   1    1   1
1 -1 500 3 5
'R'
```

9.3 Program Results

```
 G03CCF Example Program Results

  Loadings, Communalities and PSI

     0.553  -0.429   0.490   0.510
     0.568  -0.288   0.406   0.594
     0.392  -0.450   0.356   0.644
     0.740   0.273   0.623   0.377
     0.724   0.211   0.569   0.431
     0.595   0.132   0.372   0.628
```

```
Factor score coefficients

   0.193  -0.392
   0.170  -0.226
   0.109  -0.326
   0.349   0.337
   0.299   0.229
   0.169   0.098
```
