G02GNF – 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

G02GNF gives the estimate of an estimable function along with its standard error from the results from fitting a generalized linear model.

2 Specification

```
 SUBROUTINE G02GNF(IP, IRANK, B, COV, V, LDV, F, EST, STAT, SESTAT,
1                  Z, TOL, WK, IFAIL)
 INTEGER           IP, IRANK, LDV, IFAIL
 real              B(IP), COV((IP*(IP+1)/2)), V(LDV,IP+7), F(IP),
1                  STAT, SESTAT, Z, TOL, WK(IP)
 LOGICAL           EST
```

3 Description

This routine computes the estimates of an estimable function for a generalized linear model which is not of full rank. It is intended for use after a call to G02GAF, G02GBF, G02GCF or G02GDF. An estimable function is a linear combination of the parameters such that it has a unique estimate. For a full rank model all linear combinations of parameters are estimable.

In the case of a model not of full rank the routines use a singular value decomposition (SVD) to find the parameter estimates, $\hat{\beta}$, and their variance-covariance matrix. Given the upper triangular matrix R obtained from the QR decomposition of the independent variables the SVD gives:

$$R = Q_* \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} P^T,$$

where D is a k by k diagonal matrix with non-zero diagonal elements, k being the rank of R, and Q_* and P are p by p orthogonal matrices. This leads to a solution:

$$\hat{\beta} = P_1 D^{-1} Q_{*_1}^T c_1$$

P_1 being the first k columns of P, i.e., $P = (P_1 P_0)$; Q_{*_1} being the first k columns of Q_* and c_1 being the first p elements of c.

Details of the SVD, are made available, in the form of the matrix P^*:

$$P^* = \begin{pmatrix} D^{-1} P_1^T \\ P_0^T \end{pmatrix}$$

as described by G02GAF, G02GBF, G02GCF and G02GDF.

A linear function of the parameters, $F = f^T \beta$, can be tested to see if it is estimable by computing $\zeta = P_0^T f$. If ζ is zero, then the function is estimable, if not, the function is not estimable. In practice $|\zeta|$ is tested against some small quantity η.

Given that F is estimable it can be estimated by $f^T \hat{\beta}$ and its standard error calculated from the variance-covariance matrix of $\hat{\beta}$, C_β, as

$$\text{se}(F) = \sqrt{f^T C_\beta f}$$

Also a z statistic:

$$z = \frac{f^T \hat{\beta}}{\text{se}(F)},$$

can be computed. The distribution of z will be approximately Normal.

4 References

[1] Golub G H and van Loan C F (1996) *Matrix Computations* Johns Hopkins University Press (3rd Edition), Baltimore

[2] McCullagh P and Nelder J A (1983) *Generalized Linear Models* Chapman and Hall

[3] Searle S R (1971) *Linear Models* Wiley

5 Parameters

1: IP — INTEGER *Input*

On entry: the number of terms in the linear model, p.

Constraint: IP ≥ 1.

2: IRANK — INTEGER *Input*

On entry: the rank of the dependent variables, k.

Constraint: $1 \leq$ IRANK $\leq$ IP.

3: B(IP) — ***real*** array *Input*

On entry: the IP values of the estimates of the parameters of the model, $\hat{\beta}$.

4: COV((IP∗(IP+1)/2)) — ***real*** array *Input*

On entry: the upper triangular part of the variance-covariance matrix of the IP parameter estimates given in B. They are stored packed by column, i.e., the covariance between the parameter estimate given in B(i) and the parameter estimate given in B(j), $j \geq i$, is stored in COV($j \times (j-1)/2 + i$).

5: V(LDV,IP+7) — ***real*** array *Input*

On entry: V as returned by G02GAF, G02GBF, G02GCF and G02GDF.

6: LDV — INTEGER *Input*

On entry: the first dimension of the array V as declared in the (sub)program from which G02GNF is called.

Constraint: LDV $\geq$ IP.

7: F(IP) — ***real*** array *Input*

On entry: the linear function to be estimated, f.

8: EST — LOGICAL *Output*

On exit: EST indicates if the function was estimable.

If EST = .TRUE., then the function is estimable.
If EST = .FALSE., then the function is not estimable and STAT, SESTAT and Z are not set.

9: STAT — ***real*** *Output*

On exit: if EST = .TRUE., STAT contains the estimate of the function, $f^T\hat{\beta}$

10: SESTAT — ***real*** *Output*

On exit: if EST = .TRUE., SESTAT contains the standard error of the estimate of the function, se(F).

11: Z — ***real*** *Output*

On exit: if EST = .TRUE., Z contains the z statistic for the test of the function being equal to zero.

12: TOL — ***real*** *Input*

On entry: the tolerance value used in the check for estimability, η.

If TOL ≤ 0.0, then $\sqrt{\epsilon}$, where ϵ is the ***machine precision***, is used instead.

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

14: 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 or gives a warning (see Section 6).

For this routine, because the values of output parameters may be useful even if IFAIL $\neq 0$ on exit, users are recommended to set IFAIL to -1 before entry. **It is then 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 specified by the routine:

IFAIL = 1

On entry, IP < 1,
or IRANK < 1,
or IRANK $>$ IP,
or LDV $<$ IP,

IFAIL = 2

On entry, IRANK = IP. In this case EST is returned as true and all statistics are calculated.

IFAIL = 3

Standard error of statistic = 0.0, this may be due to rounding errors if the standard error is very small or due to miss-specified inputs COV and F.

7 Accuracy

The computations are believed to be stable.

8 Further Comments

The value of estimable functions is independent of the solution chosen from the many possible solutions. While G02GNF may be used to estimate functions of the parameters of the model as computed by G02GKF, β_c, these must be expressed in terms of the original parameters, β. The relation between the two sets of parameters may not be straightforward.

9 Example

A loglinear model is fitted to a 3 by 5 contingency table by G02GCF. The model consists of terms for rows and columns. The table is:

$$\begin{array}{rrrrr} 141 & 67 & 114 & 79 & 39 \\ 131 & 66 & 143 & 72 & 35 \\ 36 & 14 & 38 & 28 & 16 \end{array} .$$

The number of functions to be tested is read in, then the linear functions themselves are read in and tested with G02GNF. The results of G02GNF are printed.

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.

```
*     G02GNF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NMAX, MMAX
      PARAMETER        (NMAX=15,MMAX=9)
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real             A, DEV, EPS, SESTAT, STAT, TOL, Z
      INTEGER          I, IDF, IFAIL, IP, IPRINT, IRANK, J, M, MAXIT, N,
     +                 NESTFN
      LOGICAL          EST
*     .. Local Arrays ..
      real             B(MMAX), COV((MMAX*MMAX+MMAX)/2), F(MMAX),
     +                 SE(MMAX), V(NMAX,7+MMAX),
     +                 WK((MMAX*MMAX+3*MMAX+22)/2), WT(NMAX),
     +                 X(NMAX,MMAX), Y(NMAX)
      INTEGER          ISX(MMAX)
*     .. External Subroutines ..
      EXTERNAL         G02GCF, G02GNF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G02GNF Example Program Results'
*     Skip heading in data file
      READ (NIN,*)
      READ (NIN,*) N, M, IPRINT
      IF (N.LE.NMAX .AND. M.LT.MMAX) THEN
         DO 20 I = 1, N
            READ (NIN,*) (X(I,J),J=1,M), Y(I)
   20    CONTINUE
         READ (NIN,*) (ISX(J),J=1,M), IP
*        Set control parameters
         EPS = 0.000001e0
         TOL = 0.00005e0
         MAXIT = 10
         IFAIL = -1
*
*        Fit Log-linear model using G02GCF
         CALL G02GCF('L','M','N','U',N,X,NMAX,M,ISX,IP,Y,WT,A,DEV,IDF,B,
     +              IRANK,SE,COV,V,NMAX,TOL,MAXIT,IPRINT,EPS,WK,IFAIL)
*
         IF (IFAIL.EQ.0 .OR. IFAIL.GE.7) THEN
            WRITE (NOUT,*)
            WRITE (NOUT,99999) 'Deviance = ', DEV
            WRITE (NOUT,99998) 'Degrees of freedom = ', IDF
            WRITE (NOUT,*)
            WRITE (NOUT,*) '      Estimate     Standard error'
            WRITE (NOUT,*)
            DO 40 I = 1, IP
               WRITE (NOUT,99997) B(I), SE(I)
   40       CONTINUE
            READ (NIN,*) NESTFN
            DO 60 I = 1, NESTFN
               READ (NIN,*) (F(J),J=1,IP)
               IFAIL = -1
```

```
*
                CALL G02GNF(IP,IRANK,B,COV,V,NMAX,F,EST,STAT,SESTAT,Z,
     +                      TOL,WK,IFAIL)
*
                IF (IFAIL.EQ.0 .OR. IFAIL.EQ.2) THEN
                   WRITE (NOUT,*)
                   WRITE (NOUT,99996) 'Function ', I
                   WRITE (NOUT,99995) (F(J),J=1,IP)
                   WRITE (NOUT,*)
                   IF (EST) THEN
                      WRITE (NOUT,99994) 'STAT = ', STAT, ' SE = ',
     +                  SESTAT, ' Z = ', Z
                   ELSE
                      WRITE (NOUT,*) 'Function not estimable'
                   END IF
                END IF
   60        CONTINUE
          END IF
        END IF
        STOP
*
99999 FORMAT (1X,A,e12.4)
99998 FORMAT (1X,A,I2)
99997 FORMAT (1X,2F14.4)
99996 FORMAT (1X,A,I4)
99995 FORMAT (1X,5F8.2)
99994 FORMAT (1X,A,F10.4,A,F10.4,A,F10.4)
      END
```

9.2 Program Data

```
G02GNF Example Program Data
 15 8 0
1.0 0.0 0.0 1.0 0.0 0.0 0.0 0.0 141.
1.0 0.0 0.0 0.0 1.0 0.0 0.0 0.0  67.
1.0 0.0 0.0 0.0 0.0 1.0 0.0 0.0 114.
1.0 0.0 0.0 0.0 0.0 0.0 1.0 0.0  79.
1.0 0.0 0.0 0.0 0.0 0.0 0.0 1.0  39.
0.0 1.0 0.0 1.0 0.0 0.0 0.0 0.0 131.
0.0 1.0 0.0 0.0 1.0 0.0 0.0 0.0  66.
0.0 1.0 0.0 0.0 0.0 1.0 0.0 0.0 143.
0.0 1.0 0.0 0.0 0.0 0.0 1.0 0.0  72.
0.0 1.0 0.0 0.0 0.0 0.0 0.0 1.0  35.
0.0 0.0 1.0 1.0 0.0 0.0 0.0 0.0  36.
0.0 0.0 1.0 0.0 1.0 0.0 0.0 0.0  14.
0.0 0.0 1.0 0.0 0.0 1.0 0.0 0.0  38.
0.0 0.0 1.0 0.0 0.0 0.0 1.0 0.0  28.
0.0 0.0 1.0 0.0 0.0 0.0 0.0 1.0  16.
 1   1   1   1   1   1   1   1   9
 3
 1.0  1.0  0.0 0.0  1.0 0.0 0.0 0.0 0.0
 0.0  1.0 -1.0 0.0  0.0 0.0 0.0 0.0 0.0
 0.0  1.0  0.0 0.0  0.0 0.0 0.0 0.0 0.0
```

9.3 Program Results

```
G02GNF Example Program Results

Deviance =   0.9038E+01
Degrees of freedom =  8

      Estimate     Standard error

        2.5977         0.0258
        1.2619         0.0438
        1.2777         0.0436
        0.0580         0.0668
        1.0307         0.0551
        0.2910         0.0732
        0.9876         0.0559
        0.4880         0.0675
       -0.1996         0.0904

Function    1
    1.00    1.00    0.00    0.00    1.00
    0.00    0.00    0.00    0.00

STAT =     4.8903 SE =     0.0674 Z =    72.5934

Function    2
    0.00    1.00   -1.00    0.00    0.00
    0.00    0.00    0.00    0.00

STAT =    -0.0158 SE =     0.0672 Z =    -0.2350

Function    3
    0.00    1.00    0.00    0.00    0.00
    0.00    0.00    0.00    0.00

Function not estimable
```