NAG Fortran Library Routine Document

G01DCF

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

G01DCF computes an approximation to the variance-covariance matrix of an ordered set of independent observations from a Normal distribution with mean 0.0 and standard deviation 1.0.

2 Specification

```
SUBROUTINE G01DCF(N, EXP1, EXP2, SUMSSQ, VEC, IFAIL)
INTEGER            N, IFAIL
real               EXP1, EXP2, SUMSSQ, VEC(N*(N+1)/2)
```

3 Description

This routine is an adaptation of the Applied Statistics Algorithm AS 128, see Davis and Stephens (1978). An approximation to the variance-covariance matrix, V, using a Taylor series expansion of the Normal distribution function is discussed in David and Johnson (1954).

However, convergence is slow for extreme variances and covariances. The present routine uses the David–Johnson approximation to provide an initial approximation and improves upon it by use of the following identities for the matrix.

For a sample of size n, let m_i be the expected value of the ith largest order statistic, then:

(a) for any $i = 1, 2, \ldots, n$, $\sum_{j=1}^{n} V_{ij} = 1$

(b) $V_{12} = V_{11} + m_n^2 - m_n m_{n-1} - 1$

(c) the trace of V is $tr(V) = n - \sum_{i=1}^{n} m_i^2$

(d) $V_{ij} = V_{ji} = V_{rs} = V_{sr}$ where $r = n + 1 - i$, $s = n + 1 - j$ and $i, j = 1, 2, \ldots, n$. Note that only the upper triangle of the matrix is calculated and returned column-wise in vector form.

4 References

Davis C S and Stephens M A (1978) Algorithm AS 128: Approximating the covariance matrix of Normal order statistics *Appl. Statist.* **27** 206–212

David F N and Johnson N L (1954) Statistical treatment of censored data, Part 1. Fundamental formulae *Biometrika* **41** 228–240

5 Parameters

1: N – INTEGER *Input*

On entry: the sample size, n.

Constraint: N ≥ 1.

2: EXP1 – ***real*** *Input*

On entry: the expected value of the largest Normal order statistic, m_n, from a sample of size n.

3: EXP2 – ***real*** *Input*

On entry: the expected value of the second largest Normal order statistic, m_{n-1}, from a sample of size n.

4: SUMSSQ – ***real*** *Input*

On entry: the sum of squares of the expected values of the Normal order statistics from a sample of size n.

5: VEC(N*(N+1)/2) – ***real*** array *Output*

On exit: the upper triangle of the n by n variance-covariance matrix packed by column. Thus element V_{ij} is stored in VEC$(i + j \times (j-1)/2)$, for $1 \leq i \leq j \leq n$.

6: 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 $= 1$

On entry, N < 1.

7 Accuracy

For $n \leq 20$, where comparison with the exact values can be made, the maximum error is less than 0.0001.

8 Further Comments

The time taken by the routine is approximately proportional to n^2.

The arguments EXP1 ($= m_n$), EXP2 ($= m_n - 1$) and SUMSSQ ($= \sum_{j=1}^{n} m_j^2$) may be found from the expected values of the Normal order statistics obtained from G01DAF (exact) or G01DBF (approximate).

9 Example

A program to compute the variance-covariance matrix for a sample of size 6. G01DAF is called to provide values for EXP1, EXP2 and SUMSSQ.

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.

```
*     G01DCF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          N, IW
      PARAMETER        (N=6,IW=3*N/2)
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
*     .. Local Scalars ..
      real             ERREST, ETOL, EXP1, EXP2, SUMSSQ
      INTEGER          I, IFAIL, J, K
*     .. Local Arrays ..
      real             PP(N), VEC(N*(N+1)/2), WORK(IW)
*     .. External Subroutines ..
      EXTERNAL         G01DAF, G01DCF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G01DCF Example Program Results'
      ETOL = 0.0001e0
      IFAIL = 0
*
      CALL G01DAF(N,PP,ETOL,ERREST,WORK,IW,IFAIL)
*
      EXP1 = PP(N)
      EXP2 = PP(N-1)
      SUMSSQ = 0.0e0
      DO 20 I = 1, N
         SUMSSQ = SUMSSQ + PP(I)*PP(I)
   20 CONTINUE
      IFAIL = 0
*
      CALL G01DCF(N,EXP1,EXP2,SUMSSQ,VEC,IFAIL)
*
      WRITE (NOUT,*)
      WRITE (NOUT,99999) 'Sample size = ', N
      WRITE (NOUT,*) 'Variance-covariance matrix'
      K = 1
      DO 40 J = 1, N
         WRITE (NOUT,99998) (VEC(I),I=K,K+J-1)
         K = K + J
   40 CONTINUE
      STOP
*
99999 FORMAT (1X,A,I2)
99998 FORMAT (1X,6F8.4)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G01DCF Example Program Results

 Sample size =  6
 Variance-covariance matrix
   0.4159
   0.2085  0.2796
   0.1394  0.1889  0.2462
   0.1025  0.1397  0.1834  0.2462
   0.0774  0.1060  0.1397  0.1889  0.2796
   0.0563  0.0774  0.1025  0.1394  0.2085  0.4159
```