NAG Fortran Library Routine Document

G01DAF

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

G01DAF computes a set of Normal scores, i.e., the expected values of an ordered set of independent observations from a Normal distribution with mean 0.0 and standard deviation 1.0.

2 Specification

```
SUBROUTINE G01DAF(N, PP, ETOL, ERREST, WORK, IW, IFAIL)
INTEGER           N, IW, IFAIL
real              PP(N), ETOL, ERREST, WORK(IW)
```

3 Description

If a sample of n observations from any distribution (which may be denoted by $x_1, x_2, \ldots, x_n$), is sorted into ascending order, the rth smallest value in the sample is often referred to as the rth '**order statistic**', sometimes denoted by $x_{(r)}$ (see Kendall and Stuart (1969)).

The order statistics therefore have the property

$$x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}.$$

(If $n = 2r + 1$, x_{r+1} is the sample median.)

For samples originating from a known distribution, the distribution of each order statistic in a sample of given size may be determined. In particular, the expected values of the order statistics may be found by integration. If the sample arises from a Normal distribution, the expected values of the order statistics are referred to as the '**Normal scores**'. The Normal scores provide a set of reference values against which the order statistics of an actual data sample of the same size may be compared, to provide an indication of Normality for the sample (see G01AHF). Normal scores have other applications; for instance, they are sometimes used as alternatives to ranks in nonparametric testing procedures.

G01DAF computes the rth Normal score for a given sample size n as

$$E(x_{(r)}) = \int_{-\infty}^{\infty} x_r dG_r$$

where

$$dG_r = \frac{A_r^{r-1}(1 - A_r)^{n-r} dA_r}{\beta(r, n - r + 1)}, \quad A_r = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_r} e^{-t^2/2}\, dt, \quad r = 1, 2, \ldots, n,$$

and β denotes the complete Beta function.

The routine attempts to evaluate the scores so that the estimated error in each score is less than the value ETOL specified by the user. All integrations are performed in parallel and arranged so as to give good speed and reasonable accuracy.

4 References

Kendall M G and Stuart A (1969) *The Advanced Theory of Statistics (Volume 1)* (3rd Edition) Griffin

5 Parameters

1: N – INTEGER *Input*

On entry: the size of the set, n.

Constraint: N > 0.

2: PP(N) – ***real*** array *Output*

On exit: the Normal scores. PP(i) contains the value E$(x_{(i)})$, for $i = 1, 2, \ldots, n$.

3: ETOL – ***real*** *Input*

On entry: the maximum value for the estimated absolute error in the computed scores.

Constraint: ETOL > 0.0.

4: ERREST – ***real*** *Output*

On exit: a computed estimate of the maximum error in the computed scores (see Section 7).

5: WORK(IW) – ***real*** array *Workspace*

6: IW – INTEGER *Input*

On entry: the dimension of the array WORK as declared in the (sub)program from which G01DAF is called.

Constraints:

if N is even IW $\geq 3 \times$ N$/2$,
if N is odd IW $\geq 3 \times$ (N $-$ 1)$/2$.

7: 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.

IFAIL $= 2$

On entry, ETOL ≤ 0.0.

IFAIL $= 3$

The routine was unable to estimate the scores with estimated error less than ETOL. The best result obtained is returned together with the associated value of ERREST.

IFAIL = 4

On entry, if N is even, IW $< 3 \times \mathrm{N}/2$,
or if N is odd, IW $< 3 \times (\mathrm{N}-1)/2$.

7 Accuracy

Errors are introduced by evaluation of the functions dG_r and errors in the numerical integration process. Errors are also introduced by the approximation of the true infinite range of integration by a finite range $[a, b]$ but a and b are chosen so that this effect is of lower order than that of the other two factors. In order to estimate the maximum error the functions dG_r are also integrated over the range $[a, b]$. The routine returns the estimated maximum error as

$$\mathrm{ERREST} = \max_r \left[\max(|a|, |b|) \times \left| \int_a^b dG_r - 1.0 \right| \right].$$

8 Further Comments

The time taken by the routine depends on ETOL and N. For a given value of ETOL the timing varies approximately linearly with N.

9 Example

The program below generates the Normal scores for samples of size 5, 10, 15, and prints the scores and the computed error estimates.

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.

```
*     G01DAF Example Program Text
*     Mark 14 Revised.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NMAX, IW
      PARAMETER        (NMAX=15,IW=3*NMAX/2)
      INTEGER          NOUT
      PARAMETER        (NOUT=6)
*     .. Local Scalars ..
      real             ERREST, ETOL
      INTEGER          I, IFAIL, J, N
*     .. Local Arrays ..
      real             PP(NMAX), WORK(IW)
*     .. External Subroutines ..
      EXTERNAL         G01DAF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G01DAF Example Program Results'
      ETOL = 0.001e0
      DO 20 J = 5, NMAX, 5
         N = J
         IFAIL = 1
*
         CALL G01DAF(N,PP,ETOL,ERREST,WORK,IW,IFAIL)
*
         WRITE (NOUT,*)
         WRITE (NOUT,99999) 'Set size = ', N
         WRITE (NOUT,99998) 'Error tolerance (input) = ', ETOL
         WRITE (NOUT,99998) 'Error estimate (output) = ', ERREST
         WRITE (NOUT,*) 'Normal scores'
         WRITE (NOUT,99997) (PP(I),I=1,N)
         IF (IFAIL.NE.0) WRITE (NOUT,99999)
     +       'G01DAF fails with IFAIL = ', IFAIL
   20 CONTINUE
      STOP
```

```
*
99999 FORMAT (1X,A,I2)
99998 FORMAT (1X,A,e13.3)
99997 FORMAT (10X,5F10.3)
      END
```

9.2 Program Data

None.

9.3 Program Results

```
 G01DAF Example Program Results

 Set size =  5
 Error tolerance (input) =     0.100E-02
 Error estimate (output) =     0.908E-08
 Normal scores
              -1.163    -0.495     0.000     0.495     1.163

 Set size = 10
 Error tolerance (input) =     0.100E-02
 Error estimate (output) =     0.148E-07
 Normal scores
              -1.539    -1.001    -0.656    -0.376    -0.123
               0.123     0.376     0.656     1.001     1.539

 Set size = 15
 Error tolerance (input) =     0.100E-02
 Error estimate (output) =     0.222E-07
 Normal scores
              -1.736    -1.248    -0.948    -0.715    -0.516
              -0.335    -0.165     0.000     0.165     0.335
               0.516     0.715     0.948     1.248     1.736
```