NAG 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 (KIND=nag_wp) 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

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

(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

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

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 you. 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 Arguments

1: N – INTEGER *Input*

   On entry: n, the size of the set.

   Constraint: N > 0.

2: PP(N) – REAL (KIND=nag_wp) array *Output*

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

3: ETOL – REAL (KIND=nag_wp) *Input*

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

   Constraint: ETOL > 0.0.

4: ERREST – REAL (KIND=nag_wp) *Output*

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

5: WORK(IW) – REAL (KIND=nag_wp) 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 must be even, IW $\geq 3 \times$ N/2;
   > otherwise IW $\geq 3 \times$ (N − 1)/2.

7: IFAIL – INTEGER *Input/Output*

   On entry: IFAIL must be set to 0, −1 or 1. If you are unfamiliar with this argument you should refer to Section 3.4 in How to Use the NAG Library and its Documentation for details.

   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, if you are not familiar with this argument, the recommended value is 0. **When the value −1 or 1 is used it is essential to test the value of IFAIL on exit.**

   On exit: IFAIL = 0 unless the routine detects an error or a warning has been flagged (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 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, $\text{IW} < 3 \times \text{N}/2$;
or   if N is odd, $\text{IW} < 3 \times (\text{N} - 1)/2$.

IFAIL $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

See Section 3.9 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -399$

Your licence key may have expired or may not have been installed correctly.

See Section 3.8 in How to Use the NAG Library and its Documentation for further information.

IFAIL $= -999$

Dynamic memory allocation failed.

See Section 3.7 in How to Use the NAG Library and its Documentation for further information.

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 $[a, b]$. G01DAF returns the estimated maximum error as

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

8   Parallelism and Performance

G01DAF is not threaded in any implementation.

9   Further Comments

The time taken by G01DAF depends on ETOL and N. For a given value of ETOL the timing varies approximately linearly with N.

10   Example

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

10.1 Program Text

```
    Program g01dafe

!     G01DAF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g01daf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter                 :: nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)                 :: errest, etol
```

```
      Integer                              :: ifail, iw, n
!       .. Local Arrays ..
      Real (Kind=nag_wp), Allocatable  :: pp(:), work(:)
!       .. Executable Statements ..
      Write (nout,*) 'G01DAF Example Program Results'
      Write (nout,*)

!       Set the problem size
      n = 15
      etol = 0.001E0_nag_wp

      iw = 3*n/2
      Allocate (pp(n),work(iw))

!       Compute the normal scores
      ifail = 0
      Call g01daf(n,pp,etol,errest,work,iw,ifail)

!       Display results
      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(1:n)

99999 Format (1X,A,I2)
99998 Format (1X,A,E13.3)
99997 Format (10X,5F10.3)
    End Program g01dafe
```
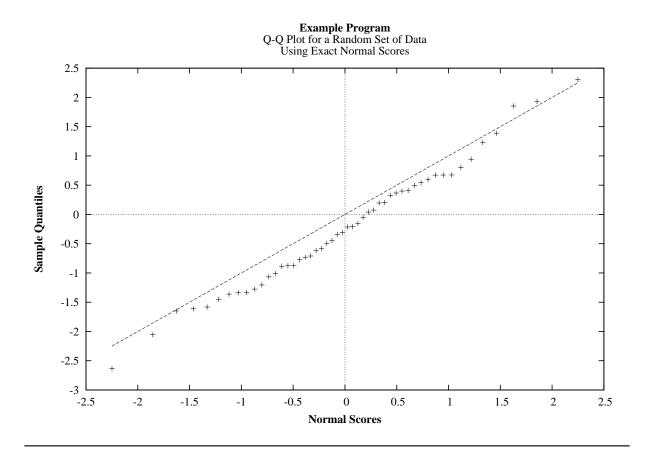
10.2 Program Data

None.

10.3 Program Results

```
 G01DAF Example Program Results

 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
```

This shows a Q-Q plot for a randomly generated set of data. The normal scores have been calculated using G01DAF and the sample quantiles obtained by sorting the observed data using M01CAF. A reference line at $y = x$ is also shown.

Example Program
Q-Q Plot for a Random Set of Data
Using Exact Normal Scores