NAG Library Routine Document

G01EYF

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

G01EYF returns the upper tail probability associated with the one sample Kolmogorov–Smirnov distribution, via the routine name.

2 Specification

```
FUNCTION G01EYF (N, D, IFAIL)
REAL (KIND=nag_wp) G01EYF
INTEGER            N, IFAIL
REAL (KIND=nag_wp) D
```

3 Description

Let $S_n(x)$ be the sample cumulative distribution function and $F_0(x)$ the hypothesised theoretical distribution function.

G01EYF returns the upper tail probability, p, associated with the one-sided Kolmogorov–Smirnov test statistic D_n^+ or D_n^-, where these one-sided statistics are defined as follows;

$$D_n^+ = \sup_x [S_n(x) - F_0(x)],$$

$$D_n^- = \sup_x [F_0(x) - S_n(x)[.$$

If $n \leq 100$ an exact method is used; for the details see Conover (1980). Otherwise a large sample approximation derived by Smirnov is used; see Feller (1948), Kendall and Stuart (1973) or Smirnov (1948).

4 References

Conover W J (1980) *Practical Nonparametric Statistics* Wiley

Feller W (1948) On the Kolmogorov–Smirnov limit theorems for empirical distributions *Ann. Math. Statist.* **19** 179–181

Kendall M G and Stuart A (1973) *The Advanced Theory of Statistics (Volume 2)* (3rd Edition) Griffin

Siegel S (1956) *Non-parametric Statistics for the Behavioral Sciences* McGraw–Hill

Smirnov N (1948) Table for estimating the goodness of fit of empirical distributions *Ann. Math. Statist.* **19** 279–281

5 Arguments

1: N – INTEGER *Input*

On entry: n, the number of observations in the sample.

Constraint: N $\geq$ 1.

2: D – REAL (KIND=nag_wp) *Input*

On entry: contains the test statistic, D_n^+ or D_n^-.

Constraint: 0.0 $\leq$ D $\leq$ 1.0.

3: 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, D < 0.0,
or D > 1.0.

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

The large sample distribution used as an approximation to the exact distribution should have a relative error of less than 2.5% for most cases.

8 Parallelism and Performance

G01EYF is not threaded in any implementation.

9 Further Comments

The upper tail probability for the two-sided statistic, $D_n = \max\left(D_n^+, D_n^-\right)$, can be approximated by twice the probability returned via G01EYF, that is $2p$. (Note that if the probability from G01EYF is greater than 0.5 then the two-sided probability should be truncated to 1.0). This approximation to the

tail probability for D_n is good for small probabilities, (e.g., $p \leq 0.10$) but becomes very poor for larger probabilities.

The time taken by the routine increases with n, until $n > 100$. At this point the approximation is used and the time decreases significantly. The time then increases again modestly with n.

10 Example

The following example reads in 10 different sample sizes and values for the test statistic D_n. The upper tail probability is computed and printed for each case.

10.1 Program Text

```
    Program g01eyfe

!     G01EYF Example Program Text

!     Mark 26 Release. NAG Copyright 2016.

!     .. Use Statements ..
      Use nag_library, Only: g01eyf, nag_wp
!     .. Implicit None Statement ..
      Implicit None
!     .. Parameters ..
      Integer, Parameter               :: nin = 5, nout = 6
!     .. Local Scalars ..
      Real (Kind=nag_wp)               :: d, prob
      Integer                          :: ifail, n
!     .. Executable Statements ..
      Write (nout,*) 'G01EYF Example Program Results'
      Write (nout,*)

!     Skip heading in data file
      Read (nin,*)

!     Display titles
      Write (nout,*) '    D       N    One-sided probability'
      Write (nout,*)

d_lp: Do
        Read (nin,*,Iostat=ifail) n, d
        If (ifail/=0) Then
          Exit d_lp
        End If

!       Calculate probability
        ifail = 0
        prob = g01eyf(n,d,ifail)

!       Display results
        Write (nout,99999) d, n, prob
      End Do d_lp

99999 Format (1X,F7.4,2X,I4,10X,F7.4)
    End Program g01eyfe
```

10.2 Program Data

```
G01EYF Example Program Data
10  0.323
10  0.369
10  0.409
10  0.457
10  0.489
```

```
400 0.0535
400 0.061
400 0.068
400 0.076
400 0.0815
```

10.3 Program Results

```
 G01EYF Example Program Results

     D       N    One-sided probability

  0.3230    10           0.0994
  0.3690    10           0.0497
  0.4090    10           0.0251
  0.4570    10           0.0099
  0.4890    10           0.0050
  0.0535   400           0.1001
  0.0610   400           0.0502
  0.0680   400           0.0243
  0.0760   400           0.0096
  0.0815   400           0.0048
```