NAG Fortran 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

```
real FUNCTION G01EYF(N, D, IFAIL)
INTEGER              N, IFAIL
real                 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;

$$\mathrm{D}_n^+ = \sup_x[S_n(x) - F_0(x)],$$

$$\mathrm{D}_n^- = \sup_x[F_0(x) - S_n(x)].$$

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

Siegel S (1956) *Nonparametric Statistics for the Behavioral Sciences* McGraw-Hill

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

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

5 Parameters

1: N – INTEGER *Input*

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

Constraint: $\mathrm{N} \geq 1$.

2: D – ***real*** *Input*

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

Constraint: $0.0 \leq \mathrm{D} \leq 1.0$.

3: 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, D < 0.0
or D > 1.0.

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 Further Comments

The upper tail probability for the two-sided statistic, $D_n = \max(\mathrm{D}_n^+, \mathrm{D}_n^-)$, 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.

9 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.

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.

```
*     G01EYF Example Program Text
*     Mark 14 Release.  NAG Copyright 1989.
*     .. Parameters ..
      INTEGER          NIN, NOUT
      PARAMETER        (NIN=5,NOUT=6)
*     .. Local Scalars ..
      real             D, PROB
      INTEGER          IFAIL, N
*     .. External Functions ..
      real             G01EYF
```

```
      EXTERNAL         G01EYF
*     .. Executable Statements ..
      WRITE (NOUT,*) 'G01EYF Example Program Results'
      WRITE (NOUT,*)
      WRITE (NOUT,*) '    D       N    One-sided probability'
      WRITE (NOUT,*)
*     Skip heading in data file
      READ (NIN,*)
   20 READ (NIN,*,END=40) N, D
      IFAIL = 0
*
      PROB = G01EYF(N,D,IFAIL)
*
      WRITE (NOUT,99999) D, N, PROB
      GO TO 20
   40 STOP
*
99999 FORMAT (1X,F7.4,2X,I4,10X,F7.4)
      END
```

9.2 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
```

9.3 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
```