NAG Toolbox for MATLAB

g01ey

1 Purpose

g01ey returns the upper tail probability associated with the one sample Kolmogorov–Smirnov distribution, via the function name.

2 Syntax

```
[result, ifail] = g01ey(n, d)
```

3 Description

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

g01ey 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 Parameters

5.1 Compulsory Input Parameters

1: **n – int32 scalar**

n, the number of observations in the sample.

Constraint: $\mathbf{n} \geq 1$.

2: **d – double scalar**

Contains the test statistic, D_n^+ or D_n^-.

Constraint: $0.0 \leq \mathbf{d} \leq 1.0$.

5.2 Optional Input Parameters

None.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **result – double scalar**

The result of the function.

2: **ifail – int32 scalar**

0 unless the function detects an error (see Section 6).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

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, $D_n = \max\left(D_n^+, D_n^-\right)$, can be approximated by twice the probability returned via g01ey, that is $2p$. (Note that if the probability from g01ey 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 function 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

```
n = int32(10);
d = 0.323;
[result, ifail] = g01ey(n, d)
```

```
result =
    0.0994
ifail =
           0
```