NAG Toolbox for MATLAB

g01da

1 Purpose

g01da 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 Syntax

```
[pp, errest, ifail] = g01da(n, etol)
```

3 Description

If a sample of n observations from any distribution (which may be denoted by $x_1, x_2, \dots, x_n$), 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

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

(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 g01ah). Normal scores have other applications; for instance, they are sometimes used as alternatives to ranks in nonparametric testing procedures.

g01da computes the rth Normal score for a given sample size n as

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

where

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

and β denotes the complete Beta function.

The function 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 Parameters

5.1 Compulsory Input Parameters

1: **n – int32 scalar**

n, the size of the set.

Constraint: $\mathbf{n} > 0$.

2: **etol – double scalar**

The maximum value for the estimated absolute error in the computed scores.

Constraint: **etol** > 0.0.

5.2 Optional Input Parameters

None.

5.3 Input Parameters Omitted from the MATLAB Interface

work, iw

5.4 Output Parameters

1: **pp(n) – double array**

The Normal scores. **pp**(i) contains the value $E(x_{(i)})$, for $i = 1, 2, \ldots, n$.

2: **errest – double scalar**

A computed estimate of the maximum error in the computed scores (see Section 7).

3: **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, **etol** ≤ 0.0.

ifail $= 3$

The function 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** $< 3 \times$ **n**$/2$;
or if **n** is odd, **iw** $< 3 \times ($**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 $[a, b]$. g01da returns the estimated maximum error as

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

8 Further Comments

The time taken by g01da depends on **etol** and **n**. For a given value of **etol** the timing varies approximately linearly with **n**.

9 Example

```
n = int32(5);
etol = 0.001;
[pp, errest, ifail] = g01da(n, etol)
```

```
pp =
   -1.1630
   -0.4950
         0
    0.4950
    1.1630
errest =
   9.0800e-09
ifail =
           0
```