NAG Toolbox for MATLAB

g01ae

1 Purpose

g01ae constructs a frequency distribution of a variable, according to either user-supplied, or function-calculated class boundary-values.

2 Syntax

```
[cb, ifreq, xmin, xmax, ifail] = g01ae(x, iclass, cb, 'n', n, 'k', k)
```

3 Description

The data consists of a sample of n observations of a continuous variable, denoted by x_i, for $i = 1, 2, \ldots, n$. Let $a = \min(x_1, \ldots, x_n)$ and $b = \max(x_1, \ldots, x_n)$.

g01ae constructs a frequency distribution with $k(> 1)$ classes denoted by f_i, for $i = 1, 2, \ldots, k$.

The boundary-values may be either user-supplied, or function-calculated, and are denoted by y_j, for $j = 1, 2, \ldots, k - 1$.

If the boundary-values of the classes are to be function-calculated, then they are determined in one of the following ways:

(a) if $k > 2$, the range of x values is divided into $k - 2$ intervals of equal length, and two extreme intervals, defined by the class boundary-values $y_1, y_2, \ldots, y_{k-1}$;

(b) if $k = 2$, $y_1 = \frac{1}{2}(a + b)$.

However formed, the values $y_1, \ldots, y_{k-1}$ are assumed to be in ascending order. The class frequencies are formed with

$$f_1 = \text{ the number of } x \text{ values in the interval } (-\infty, y_1)$$
$$f_i = \text{ the number of } x \text{ values in the interval } [y_{i-1}, i_k), \qquad i = 2, \ldots, k - 1$$
$$f_k = \text{ the number of } x \text{ values in the interval } [y_{k-1}, \infty),$$

where [means inclusive, and) means exclusive. If the class boundary-values are function-calculated and $k > 2$, then $f_1 = f_k = 0$, and y_1 and y_{k-1} are chosen so that $y_1 < a$ and $y_{k-1} > b$.

If a frequency distribution is required for a discrete variable, then it is suggested that you supply the class boundary-values; function-calculated boundary-values may be slightly imprecise (due to the adjustment of y_1 and y_{k-1} outlined above) and cause values very close to a class boundary to be assigned to the wrong class.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **x(n) – double array**

The sample of observations of the variable for which the frequency distribution is required, x_i, for $i = 1, 2, \ldots, n$. The values may be in any order.

2: **iclass – int32 scalar**

Indicates whether class boundary-values are to be calculated within g01ae, or are supplied by you.

If **iclass** $\neq 1$, then the class boundary-values are to be calculated within the function.

If **iclass** $= 1$, they are user-supplied.

3: **cb(k) – double array**

If **iclass** $= 0$, then the elements of **cb** need not be assigned values, as g01ae calculates $k - 1$ class boundary-values.

If **iclass** $= 1$, the first $k - 1$ elements of **cb** must contain the user-supplied class boundary-values, in ascending order.

In both cases, the element **cb**(k) need not be assigned, as it is not used in the function.

Constraint: **cb**$(i) <$ **cb**$(i + 1)$ if **iclass** $= 1$, for $i = 1, 2, \ldots, k - 2$.

5.2 Optional Input Parameters

1: **n – int32 scalar**

Default: The dimension of the array **x**.

n, the number of observations.

Constraint: **n** ≥ 1.

2: **k – int32 scalar**

Default: The dimension of the arrays **cb**, **ifreq**. (An error is raised if these dimensions are not equal.)

k, the number of classes desired in the frequency distribution. Whether or not class boundary-values are user-supplied, **k** must include the two extreme classes which stretch to $\pm\infty$.

Constraint: **k** ≥ 2.

5.3 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **cb(k) – double array**

The first $k - 1$ elements of **cb** contain the class boundary-values in ascending order.

2: **ifreq(k) – int32 array**

The elements of **ifreq** contain the frequencies in each class, f_i for $i = 1, 2, \ldots, k$. In particular **ifreq**(1) contains the frequency of the class up to **cb**(1), f_1, and **ifreq**(k) contains the frequency of the class greater than **cb**$(k - 1)$, f_k.

3: **xmin – double scalar**

The smallest value in the sample, a.

4: **xmax – double scalar**

The largest value in the sample, b.

5: **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, $\mathbf{k} < 2$.

ifail $= 2$

On entry, $\mathbf{n} < 1$.

ifail $= 3$

On entry, the user-supplied class boundary-values are not in ascending order.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by g01ae increases with **k** and **n**. It also depends on the distribution of the sample observations.

9 Example

```
x = [22.3;
     21.6;
     22.6;
     22.4;
     22.4;
     22.4;
     22.1;
     21.9;
     23.1;
     23.4;
     23.4;
     22.6;
     22.5;
     22.5;
     22.1;
     22.6;
     22.3;
     22.4;
     21.8;
     22.3;
     22.1;
     23.6;
     20.8;
     22.2;
     23.1;
     21.1;
     21.7;
     21.4;
     21.6;
     22.5;
     21.2;
     22.6;
     22.2;
     22.2;
     21.4;
     21.7;
     23.2;
```

```
     23.1;
     22.3;
     22.3;
     21.1;
     21.4;
     21.5;
     21.8;
     22.8;
     21.4;
     20.7;
     21.6;
     23.2;
     23.6;
     22.7;
     21.7;
     23;
     21.9;
     22.6;
     22.1;
     22.2;
     23.4;
     21.5;
     23;
     22.8;
     21.4;
     23.2;
     21.8;
     21.2;
     22;
     22.4;
     22.8;
     23.2;
     23.6];
iclass = int32(0);
cb = zeros(7, 1);
[cbOut, ifreq, xmin, xmax, ifail] = g01ae(x, iclass, cb)
```

```
cbOut =
   20.6986
   21.2791
   21.8597
   22.4403
   23.0209
   23.6015
         0
ifreq =
           0
           6
          16
          21
          14
          13
           0
xmin =
   20.7000
xmax =
   23.6000
ifail =
           0
```