NAG Toolbox for MATLAB

g02bn

1 Purpose

g02bn computes Kendall and/or Spearman non-parametric rank correlation coefficients for a set of data; the data array is overwritten with the ranks of the observations.

2 Syntax

```
[x, rr, ifail] = g02bn(n, x, itype, 'm', m)
```

3 Description

The input data consists of n observations for each of m variables, given as an array

$$\left[x_{ij}\right], \qquad i = 1, 2, \ldots, n (n \geq 2), j = 1, 2, \ldots, m (m \geq 2),$$

where x_{ij} is the ith observation of the jth variable.

The quantities calculated are:

(a) Ranks:

For a given variable, j say, each of the n observations, $x_{1j}, x_{2j}, \ldots, x_{nj}$, has associated with it an additional number, the 'rank' of the observation, which indicates the magnitude of that observation relative to the magnitudes of the other $n - 1$ observations on that same variable.

The smallest observation for variable j is assigned the rank 1, the second smallest observation for variable j the rank 2, the third smallest the rank 3, and so on until the largest observation for variable j is given the rank n.

If a number of cases all have the same value for the given variable, j, then they are each given an 'average' rank, e.g., if in attempting to assign the rank $h + 1$, k observations were found to have the same value, then instead of giving them the ranks

$$h + 1, h + 2, \ldots, h + k,$$

all k observations would be assigned the rank

$$\frac{2h + k + 1}{2}$$

and the next value in ascending order would be assigned the rank

$$h + k + 1.$$

The process is repeated for each of the m variables.

Let y_{ij} be the rank assigned to the observation x_{ij} when the jth variable is being ranked. The actual observations x_{ij} are replaced by the ranks y_{ij}.

(b) Non parametric rank correlation coefficients

(i) Kendall's tau:

$$R_{jk} = \frac{\sum_{h=1}^{n} \sum_{i=1}^{n} \operatorname{sign}\left(y_{hj} - y_{ij}\right) \operatorname{sign}(y_{hk} - y_{ik})}{\sqrt{\left[n(n-1) - T_j\right]\left[n(n-1) - T_k\right]}}, \qquad j, k = 1, 2, \ldots, m,$$

where $\operatorname{sign} u = 1$ if $u > 0$,

$\operatorname{sign} u = 0$ if $u = 0$,

$\operatorname{sign} u = -1$ if $u < 0$, and $T_j = \sum t_j(t_j - 1)$, where t_j is the number of ties of a particular value of variable j, and the summation is over all tied values of variable j.

(ii) Spearman's:

$$R_{jk}^* = \frac{n(n^2-1) - 6\sum_{i=1}^{n}\left(y_{ij} - y_{ik}\right)^2 - \frac{1}{2}\left(T_j^* + T_k^*\right)}{\sqrt{\left[n(n^2-1) - T_j^*\right]\left[n(n^2-1) - T_k^*\right]}}, \qquad j,k = 1,2,\ldots,m,$$

where $T_j^* = \sum t_j(t_j^2 - 1)$, t_j being the number of ties of a particular value of variable j, and the summation being over all tied values of variable j.

4 References

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

5 Parameters

5.1 Compulsory Input Parameters

1: **n – int32 scalar**

n, the number of observations or cases.

Constraint: $\mathbf{n} \geq 2$.

2: **x(ldx,m) – double array**

ldx, the first dimension of the array, must be at least **n**.

x(i,j) must be set to x_{ij}, the value of the ith observation on the jth variable, for $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,m$.

3: **itype – int32 scalar**

The type of correlation coefficients which are to be calculated.

itype $= -1$

Only Kendall's tau coefficients are calculated.

itype $= 0$

Both Kendall's tau and Spearman's coefficients are calculated.

itype $= 1$

Only Spearman's coefficients are calculated.

Constraint: **itype** $= -1$, 0 or 1.

5.2 Optional Input Parameters

1: **m – int32 scalar**

Default: The dimension of the arrays **x**, **xbar**, **std**, **ssp**, **r**. (An error is raised if these dimensions are not equal.)

m, the number of variables.

Constraint: $\mathbf{m} \geq 2$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldx, ldrr, kworka, kworkb, work1, work2

5.4 Output Parameters

1: **x(ldx,m) – double array**

x(i,j) contains the rank y_{ij} of the observation x_{ij}, for $i = 1, 2, \ldots, n$ and $j = 1, 2, \ldots, m$.

2: **rr(ldrr,m) – double array**

The requested correlation coefficients.

If only Kendall's tau coefficients are requested (**itype** $= -1$), **rr**(j, k) contains Kendall's tau for the jth and kth variables.

if only Spearman's coefficients are requested (**itype** $= 1$), **rr**(j, k) contains Spearman's rank correlation coefficient for the jth and kth variables.

If both Kendall's tau and Spearman's coefficients are requested (**itype** $= 0$), the upper triangle of **rr** contains the Spearman coefficients and the lower triangle the Kendall coefficients. That is, for the jth and kth variables, where j is less than k, **rr**(j, k) contains the Spearman rank correlation coefficient, and **rr**(k, j) contains Kendall's tau, for $j, k = 1, 2, \ldots, m$.

(Diagonal terms, **rr**(j, j), are unity for all three values of **itype**.)

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** < 2.

ifail $= 2$

On entry, **m** < 2.

ifail $= 3$

On entry, **ldx** $<$ **n**,
or **ldrr** $<$ **m**.

ifail $= 4$

On entry, **itype** < -1,
or **itype** > 1.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by g02bn depends on n and m.

9 Example

```
n = int32(9);
x = [1.7, 1, 0.5;
```

```
     2.8, 4, 3;
     0.6, 6, 2.5;
     1.8, 9, 6;
     0.99, 4, 2.5;
     1.4, 2, 5.5;
     1.8, 9, 7.5;
     2.5, 7, 0;
     0.99, 5, 3];
itype = int32(0);
[xOut, rr, ifail] = g02bn(n, x, itype)
```

```
xOut =
    5.0000    1.0000    2.0000
    9.0000    3.5000    5.5000
    1.0000    6.0000    3.5000
    6.5000    8.5000    8.0000
    2.5000    3.5000    3.5000
    4.0000    2.0000    7.0000
    6.5000    8.5000    9.0000
    8.0000    7.0000    1.0000
    2.5000    5.0000    5.5000
rr =
    1.0000    0.2246    0.1186
    0.0294    1.0000    0.3814
    0.1176    0.2353    1.0000
ifail =
           0
```