NAG Toolbox for MATLAB

g01af

1 Purpose

g01af performs the analysis of a two-way $r \times c$ contingency table or classification. If $r = c = 2$, and the total number of objects classified is 40 or fewer, then the probabilities for Fisher's exact test are computed. Otherwise, a test statistic is computed (with Yates' correction when $r = c = 2$), which under the assumption of no association between the classifications has approximately a chi-square distribution with $(r - 1) \times (c - 1)$ degrees of freedom.

2 Syntax

```
[nobs, num, pred, chis, p, npos, ndf, m1, n1, ifail] = g01af(m, nobs,
num, 'n', n)
```

3 Description

The data consist of the frequencies for the two-way classification, denoted by n_{ij}, for $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$ with $m, n > 1$.

A check is made to see whether any row or column of the matrix of frequencies consists entirely of zeros, and if so, the matrix of frequencies is reduced by omitting that row or column. Suppose the final size of the matrix is m_1 by n_1 $(m_1, n_1 > 1)$, and let

$R_i = \sum_{j=1}^{n_1} n_{ij}$, the total frequency for the ith row, for $i = 1, 2, \ldots, m_1$,

$C_j = \sum_{i=1}^{m_1} n_{ij}$, the total frequency for the jth column, for $j = 1, 2, \ldots, n_1$, and

$T = \sum_{i=1}^{m_1} R_i = \sum_{j=1}^{n_1} C_j$, the total frequency.

There are two situations:

(i) If $m_1 > 2$ and/or $n_1 > 2$, or $m_1 = n_1 = 2$ and $T > 40$, then the matrix of expected frequencies, denoted by r_{ij}, for $i = 1, 2, \ldots, m_1$; $j = 1, 2, \ldots, n_1$, and the test statistic, χ^2, are computed, where

$$r_{ij} = R_i C_j / T, \qquad i = 1, 2, \ldots, m_1; j = 1, 2, \ldots, n_1$$

and

$$\chi^2 = \sum_{i=1}^{m_1} \sum_{j=1}^{n_1} \left[|r_{ij} - n_{ij}| - Y \right]^2 / r_{ij},$$

where

$$Y = \begin{cases} \frac{1}{2} & \text{if } m_1 = n_1 = 2 \\ 0 & \text{otherwise} \end{cases}$$

is Yates' correction for continuity.

Under the assumption that there is no association between the two classifications, X^2 will have approximately a chi-square distribution with $(m_1 - 1) \times (n_1 - 1)$ degrees of freedom.

An option exists which allows for further 'shrinkage' of the matrix of frequencies in the case where $r_{ij} < 1$ for the (i,j)th cell. If this is the case, then row i or column j will be combined with the adjacent row or column with smaller total. Row i is selected for combination if $R_i \times m_1 \leq C_j \times n_1$. This 'shrinking' process is continued until $r_{ij} \geq 1$ for all cells (i,j).

(ii) If $m_1 = n_1 = 2$ and $T \leq 40$, the probabilities to enable Fisher's exact test to be made are computed.

The matrix of frequencies may be rearranged so that R_1 is the smallest marginal (i.e., column and row) total, and $C_2 \geq C_1$. Under the assumption of no association between the classifications, the probability of obtaining r entries in cell $(1,1)$ is computed where

$$P_{r+1} = \frac{R_1!R_2!C_1!C_2!}{T!r!(R_1 - r)!(C_1 - r)!(T - C_1 - R_1 + r)!}, \qquad r = 0, 1, \ldots, R_1.$$

The probability of obtaining the table of given frequencies is returned. A test of the assumption against some alternative may then be made by summing the relevant values of P_r.

4 References

None.

5 Parameters

5.1 Compulsory Input Parameters

1: **m – int32 scalar**

one more than the number of rows of the frequency matrix, **nobs**, $m + 1$.

2: **nobs(ldnob,n) – int32 array**

ldnob, the first dimension of the array, must be at least **m**.

The elements **nobs**(i,j), for $i = 1, 2, \ldots, m$; $j = 1, 2, \ldots, n$, must contain the frequencies for the two-way classification.

3: **num – int32 scalar**

The value assigned to **num** must determine whether automatic 'shrinkage' is required when any $r_{ij} < 1$, as outlined in Section 3(a).

If **num** $= 0$, shrinkage is not required, and if **num** $= 1$, shrinkage is required.

5.2 Optional Input Parameters

1: **n – int32 scalar**

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

one more than the number of columns of the frequency matrix, **nobs**, $n + 1$.

5.3 Input Parameters Omitted from the MATLAB Interface

ldnob, ldpred

5.4 Output Parameters

1: **nobs(ldnob,n) – int32 array**

Contains the following information:

nobs(i,j), for $i = 1, 2, \ldots, m_1$; $j = 1, 2, \ldots, n_1$, contain the frequencies for the two-way classification after 'shrinkage' has taken place (see Section 3).

nobs$(i, n + 1)$, for $i = 1, 2, \ldots, m_1$, contain the total frequencies in the remaining rows, R_i.

nobs$(m+1,j)$, for $j=1,2,\ldots,n_1$, contain the total frequencies in the remaining columns, C_j.

nobs$(m+1,n+1)$, contains the total frequency, T.

If any ‘shrinkage’ has occurred, then all other cells contain no useful information.

2: **num – int32 scalar**

Contains the number of elements used in the array **p**. When Fisher’s exact test for a 2×2 classification is used.

If **p** is not used, that is if Fisher’s exact test is not to be used, then **num** is set to zero.

3: **pred(ldpred,n) – double array**

The elements **pred**(i,j), where $i=1,2,\ldots,$**m1** and $j=1,2,\ldots,$**n1** contain the expected frequencies, r_{ij} corresponding to the observed frequencies **nobs**(i,j), except in the case when Fisher’s exact test for a 2×2 classification is to be used, when **pred** is not used. No other elements are utilized.

4: **chis – double scalar**

The value of the test statistic, χ^2, except when Fisher’s exact test for a 2×2 classification is used in which case it is unspecified.

5: **p(21) – double array**

p is used only when Fisher’s exact test for a 2×2 classification is to be used.

The first **num** elements contain the probabilities associated with the various possible frequency tables, P_r, for $r=0,1,\ldots,R_1$, the remainder are unspecified.

6: **npos – int32 scalar**

npos is used only when Fisher’s exact test for a 2×2 classification is to be used.

The value of **npos** is the element of **p** which contains the probability associated with the given table of frequencies.

7: **ndf – int32 scalar**

The value of **ndf** gives the number of degrees of freedom for the chi-square distribution, $(m_1-1)\times(n_1-1)$; when Fisher’s exact test is used **ndf** $=1$.

8: **m1 – int32 scalar**

The number of rows of the two-way classification, after any ‘shrinkage’, m_1.

9: **n1 – int32 scalar**

The number of columns of the two-way classification, after any ‘shrinkage’, n_1.

10: **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 $=1$

The number of rows or columns of **nobs** is less than 2, possibly after shrinkage.

ifail = 2

At least one frequency is negative, or all frequencies are zero.

ifail = 4

On entry, **ldpred** < **m**,
or **ldnob** < **m**.

7 Accuracy

The method used is believed to be stable.

8 Further Comments

The time taken by g01af will increase with **m** and **n**, except when Fisher's exact test is to be used, in which case it increases with size of the marginal and total frequencies.

If, on exit, **num** > 0, or alternatively **ndf** is 1 and **nobs**(**m**, **n**) ≤ 40, the probabilities for use in Fisher's exact test for a 2×2 classification will be calculated, and not the test statistic with approximately a chi-square distribution.

9 Example

```
m = int32(3);
nobs = [int32(86), int32(51), int32(13), int32(-1078555088);
     int32(130), int32(115), int32(41), int32(832);
           int32(-1232907664),  int32(10574844),  int32(262144),  int32(-
1078555088)];
num = int32(0);
[nobsOut, numOut, pred, chis, p, npos, ndf, m1, n1, ifail] = g01af(m,
nobs, num)
```

```
nobsOut =
          86          51          13         150
         130         115          41         286
         216         166          54         436
numOut =
           0
pred =
   74.3119   57.1101   18.5780         0
  141.6881  108.8899   35.4220         0
         0         0         0         0
chis =
    6.3522
p =
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
     0
npos =
           0
ndf =
           2
m1 =
           2
n1 =
           3
ifail =
           0
```