NAG Toolbox

nag_stat_prob_hypergeom (g01bl)

1 Purpose

nag_stat_prob_hypergeom (g01bl) returns the lower tail, upper tail and point probabilities associated with a hypergeometric distribution.

2 Syntax

```
[plek, pgtk, peqk, ifail] = nag_stat_prob_hypergeom(n, l, m, k)
[plek, pgtk, peqk, ifail] = g01bl(n, l, m, k)
```

3 Description

Let X denote a random variable having a hypergeometric distribution with parameters n, l and m ($n \geq l \geq 0$, $n \geq m \geq 0$). Then

$$\text{Prob}\{X = k\} = \frac{\binom{m}{k}\binom{n-m}{l-k}}{\binom{n}{l}},$$

where $\max(0, l-(n-m)) \leq k \leq \min(l, m)$, $0 \leq l \leq n$ and $0 \leq m \leq n$.

The hypergeometric distribution may arise if in a population of size n a number m are marked. From this population a sample of size l is drawn and of these k are observed to be marked.

The mean of the distribution $= \dfrac{lm}{n}$, and the variance $= \dfrac{lm(n-l)(n-m)}{n^2(n-1)}$.

nag_stat_prob_hypergeom (g01bl) computes for given n, l, m and k the probabilities:

$$\begin{aligned} \textbf{plek} &= \text{Prob}\{X \leq k\} \\ \textbf{pgtk} &= \text{Prob}\{X > k\} \\ \textbf{peqk} &= \text{Prob}\{X = k\}. \end{aligned}$$

The method is similar to the method for the Poisson distribution described in KnÏsel (1986).

4 References

KnÏsel L (1986) Computation of the chi-square and Poisson distribution *SIAM J. Sci. Statist. Comput.* **7** 1022–1036

5 Parameters

5.1 Compulsory Input Parameters

1: **n** – INTEGER

The parameter n of the hypergeometric distribution.

Constraint: $\textbf{n} \geq 0$.

2: **l** – INTEGER

The parameter l of the hypergeometric distribution.

Constraint: $0 \leq \mathbf{l} \leq \mathbf{n}$.

3: **m** – INTEGER

The parameter m of the hypergeometric distribution.

Constraint: $0 \leq \mathbf{m} \leq \mathbf{n}$.

4: **k** – INTEGER

The integer k which defines the required probabilities.

Constraint: $\max(0, \mathbf{l} - (\mathbf{n} - \mathbf{m})) \leq \mathbf{k} \leq \min(\mathbf{l}, \mathbf{m})$.

5.2 Optional Input Parameters

None.

5.3 Output Parameters

1: **plek** – REAL (KIND=nag_wp)

The lower tail probability, Prob$\{X \leq k\}$.

2: **pgtk** – REAL (KIND=nag_wp)

The upper tail probability, Prob$\{X > k\}$.

3: **peqk** – REAL (KIND=nag_wp)

The point probability, Prob$\{X = k\}$.

4: **ifail** – INTEGER

ifail $= 0$ unless the function detects an error (see Section 5).

6 Error Indicators and Warnings

Errors or warnings detected by the function:

ifail $= 1$

On entry, $\mathbf{n} < 0$.

ifail $= 2$

On entry, $\mathbf{l} < 0$,
or $\mathbf{l} > \mathbf{n}$.

ifail $= 3$

On entry, $\mathbf{m} < 0$,
or $\mathbf{m} > \mathbf{n}$.

ifail $= 4$

On entry, $\mathbf{k} < 0$,
or $\mathbf{k} > \mathbf{l}$,
or $\mathbf{k} > \mathbf{m}$,
or $\mathbf{k} < \mathbf{l} + \mathbf{m} - \mathbf{n}$.

ifail $= 5$

On entry, **n** is too large to be represented exactly as a double number.

ifail $= 6$

On entry, the variance (see Section 3) exceeds 10^6.

ifail $= -99$

An unexpected error has been triggered by this routine. Please contact NAG.

ifail $= -399$

Your licence key may have expired or may not have been installed correctly.

ifail $= -999$

Dynamic memory allocation failed.

7 Accuracy

Results are correct to a relative accuracy of at least 10^{-6} on machines with a precision of 9 or more decimal digits, and to a relative accuracy of at least 10^{-3} on machines of lower precision (provided that the results do not underflow to zero).

8 Further Comments

The time taken by nag_stat_prob_hypergeom (g01bl) depends on the variance (see Section 3) and on k. For given variance, the time is greatest when $k \approx lm/n$ (= the mean), and is then approximately proportional to the square-root of the variance.

9 Example

This example reads values of n, l, m and k from a data file until end-of-file is reached, and prints the corresponding probabilities.

9.1 Program Text

```
function g01bl_example

fprintf('g01bl example results\n\n');

n = nag_int([10 40 155 1000]);
l = nag_int([ 2 10  35  444]);
m = nag_int([ 5  3 122  500]);
k = nag_int([ 1  2  22  220]);

fprintf('    n   l   m   k     plek      pgtk      peqk\n');
for i = 1:4
  [plek, pgtk, peqk, ifail] = ...
  g01bl(n(i), l(i), m(i), k(i));

  fprintf('%5d%4d%4d%4d%10.5f%10.5f%10.5f\n', n(i), l(i), m(i), k(i), ...
   plek, pgtk, peqk);
end
```

9.2 Program Results

```
     g01bl example results

    n    l    m    k      plek      pgtk      peqk
   10    2    5    1   0.77778   0.22222   0.55556
   40   10    3    2   0.98785   0.01215   0.13664
  155   35  122   22   0.01101   0.98899   0.00779
 1000  444  500  220   0.42429   0.57571   0.04913
```