NAG Toolbox for MATLAB

g01bl

1 Purpose

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

2 Syntax

```
[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

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

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 $= \frac{l}{mn}$, and the variance $= \frac{lm(n-l)(n-m)}{n^2(n-1)}$.

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

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

The method is similar to the method for the Poisson distribution described in Knüsel 1986.

4 References

Knüsel 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 – int32 scalar**

The parameter n of the hypergeometric distribution.

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

2: **l – int32 scalar**

The parameter l of the hypergeometric distribution.

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

3: **m – int32 scalar**

The parameter m of the hypergeometric distribution.

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

4: **k – int32 scalar**

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 Input Parameters Omitted from the MATLAB Interface

None.

5.4 Output Parameters

1: **plek – double scalar**

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

2: **pgtk – double scalar**

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

3: **peqk – double scalar**

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

4: **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, $\mathbf{n} < 0$.

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.

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 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

```
n = int32(10);
l = int32(2);
m = int32(5);
k = int32(1);
[plek, pgtk, peqk, ifail] = g01bl(n, l, m, k)
```

```
plek =
    0.7778
pgtk =
    0.2222
peqk =
    0.5556
ifail =
           0
```