

NAG Toolbox

nag_univar_ci_binomial (g07aa)

1 Purpose

nag_univar_ci_binomial (g07aa) computes a confidence interval for the argument p (the probability of a success) of a binomial distribution.

2 Syntax

```
[pl, pu, ifail] = nag_univar_ci_binomial(n, k, clevel)
[pl, pu, ifail] = g07aa(n, k, clevel)
```

3 Description

Given the number of trials, n , and the number of successes, k , this function computes a $100(1 - \alpha)\%$ confidence interval for p , the probability argument of a binomial distribution with probability function,

$$f(x) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n,$$

where α is in the interval $(0, 1)$.

Let the confidence interval be denoted by $[p_l, p_u]$.

The point estimate for p is $\hat{p} = k/n$.

The lower and upper confidence limits p_l and p_u are estimated by the solutions to the equations;

$$\sum_{x=k}^n \binom{n}{x} p_l^x (1-p_l)^{n-x} = \alpha/2,$$

$$\sum_{x=0}^k \binom{n}{x} p_u^x (1-p_u)^{n-x} = \alpha/2.$$

Three different methods are used depending on the number of trials, n , and the number of successes, k .

1. If $\max(k, n - k) < 10^6$.

The relationship between the beta and binomial distributions (see page 38 of Hastings and Peacock (1975)) is used to derive the equivalent equations,

$$p_l = \beta_{k, n-k+1, \alpha/2},$$

$$p_u = \beta_{k+1, n-k, 1-\alpha/2},$$

where $\beta_{a,b,\delta}$ is the deviate associated with the lower tail probability, δ , of the beta distribution with arguments a and b . These beta deviates are computed using nag_stat_inv_cdf_beta (g01fe).

2. If $\max(k, n - k) \geq 10^6$ and $\min(k, n - k) \leq 1000$.

The binomial variate with arguments n and p is approximated by a Poisson variate with mean np , see page 38 of Hastings and Peacock (1975).

The relationship between the Poisson and χ^2 -distributions (see page 112 of Hastings and Peacock (1975)) is used to derive the following equations;

$$p_l = \frac{1}{2n} \chi_{2k, \alpha/2}^2,$$

$$p_u = \frac{1}{2n} \chi_{2k+2, 1-\alpha/2}^2,$$

where $\chi_{\delta, \nu}^2$ is the deviate associated with the lower tail probability, δ , of the χ^2 -distribution with ν degrees of freedom.

In turn the relationship between the χ^2 -distribution and the gamma distribution (see page 70 of Hastings and Peacock (1975)) yields the following equivalent equations;

$$p_l = \frac{1}{2n} \gamma_{k, 2; \alpha/2},$$

$$p_u = \frac{1}{2n} \gamma_{k+1, 2; 1-\alpha/2},$$

where $\gamma_{\alpha, \beta; \delta}$ is the deviate associated with the lower tail probability, δ , of the gamma distribution with shape argument α and scale argument β . These deviates are computed using `nag_stat_inv_cdf_gamma (g01ff)`.

3. If $\max(k, n - k) > 10^6$ and $\min(k, n - k) > 1000$.

The binomial variate with arguments n and p is approximated by a Normal variate with mean np and variance $np(1 - p)$, see page 38 of Hastings and Peacock (1975).

The approximate lower and upper confidence limits p_l and p_u are the solutions to the equations;

$$\frac{k - np_l}{\sqrt{np_l(1 - p_l)}} = z_{1-\alpha/2},$$

$$\frac{k - np_u}{\sqrt{np_u(1 - p_u)}} = z_{\alpha/2},$$

where z_δ is the deviate associated with the lower tail probability, δ , of the standard Normal distribution. These equations are solved using a quadratic equation solver (`nag_zeros_quadratic_real (c02aj)`).

4 References

Hastings N A J and Peacock J B (1975) *Statistical Distributions* Butterworth

Snedecor G W and Cochran W G (1967) *Statistical Methods* Iowa State University Press

5 Parameters

5.1 Compulsory Input Parameters

- 1: **n** – INTEGER

n , the number of trials.

Constraint: $\mathbf{n} \geq 1$.

- 2: **k** – INTEGER

k , the number of successes.

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

3: **clevel** – REAL (KIND=nag_wp)

The confidence level, $(1 - \alpha)$, for two-sided interval estimate. For example **clevel** = 0.95 will give a 95% confidence interval.

Constraint: $0.0 < \mathbf{clevel} < 1.0$.

5.2 Optional Input Parameters

None.

5.3 Output Parameters

1: **pl** – REAL (KIND=nag_wp)

The lower limit, p_l , of the confidence interval.

2: **pu** – REAL (KIND=nag_wp)

The upper limit, p_u , of the confidence interval.

3: **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, **n** < 1,
or **k** < 0,
or **n** < **k**,
or **clevel** ≤ 0.0,
or **clevel** ≥ 1.0.

ifail = 2

When using the relationship with the gamma distribution to calculate one of the confidence limits, the series to calculate the gamma probabilities has failed to converge. Both **pl** and **pu** are set to zero. This is a very unlikely error exit and if it occurs please contact NAG.

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

For most cases using the beta deviates the results should have a relative accuracy of $\max(0.5e-12, 50.0 \times \epsilon)$ where ϵ is the *machine precision* (see nag_machine_precision (x02aj)). Thus on machines with sufficiently high precision the results should be accurate to 12 significant figures. Some accuracy may be lost when $\alpha/2$ or $1 - \alpha/2$ is very close to 0.0, which will occur if **clevel** is very close to 1.0. This should not affect the usual confidence levels used.

The approximations used when n is large are accurate to at least 3 significant digits but usually to more.

8 Further Comments

None.

9 Example

The following example program reads in the number of deaths recorded among male recipients of war pensions in a six year period following an initial questionnaire in 1956. We consider two classes, non-smokers and those who reported that they smoked pipes only. The total number of males in each class is also read in. The data is taken from page 216 of Snedecor and Cochran (1967). An estimate of the probability of a death in the six year period in each class is computed together with 95% confidence intervals for these estimates.

9.1 Program Text

```
function g07aa_example

fprintf('g07aa example results\n\n');

n = nag_int([1067, 402]);
k = nag_int([117, 54]);
clevel = 0.95;

disp(' Probability      Confidence Interval ');
for j = 1:numel(n)
    [pl, pu, ifail] = g07aa(n(j), k(j), clevel);
    p_hat = double(k(j))/double(n(j));
    fprintf('%10.4f      (%7.4f, %7.4f)\n', p_hat, pl, pu);
end
```

9.2 Program Results

```
g07aa example results

Probability      Confidence Interval
0.1097          ( 0.0915,  0.1300)
0.1343          ( 0.1025,  0.1716)
```
