NAG Library Function Document

nag_binomial_ci (g07aac)

1 Purpose

nag_binomial_ci (g07aac) computes a confidence interval for the argument p (the probability of a success) of a binomial distribution.

2 Specification

```
#include <nag.h>
#include <nagg07.h>
```

```
void nag_binomial_ci (Integer n, Integer k, double clevel, double *pl,
     double *pu, NagError *fail)
```

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,

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

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,

   $$\begin{aligned} p_l &= \beta_{k,n-k+1,\alpha/2}, \\ p_u &= \beta_{k+1,n-k,1-\alpha/2}, \end{aligned}$$

   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_deviates_beta (g01fec).

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;

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

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

where $\chi^2_{\delta,\nu}$ 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_deviates_gamma_dist (g01ffc).

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 .

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 Arguments

1: **n** – Integer *Input*

On entry: n, the number of trials.

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

2: **k** – Integer *Input*

On entry: k, the number of successes.

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

3: **clevel** – double *Input*

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

Constraint: $0.0 <$ **clevel** < 1.0.

4: **pl** – double * *Output*

On exit: the lower limit, p_l, of the confidence interval.

5: **pu** – double * *Output*

On exit: the upper limit, p_u, of the confidence interval.

6: **fail** – NagError * *Input/Output*

The NAG error argument (see Section 2.7 in How to Use the NAG Library and its Documentation).

6 Error Indicators and Warnings

NE_ALLOC_FAIL

Dynamic memory allocation failed.
See Section 2.3.1.2 in How to Use the NAG Library and its Documentation for further information.

NE_BAD_PARAM

On entry, argument ⟨*value*⟩ had an illegal value.

NE_CONVERGENCE

When using the relationship with the gamma distribution the series to calculate the gamma probabilities has failed to converge.

NE_INT

On entry, **k** $=$ ⟨*value*⟩.
Constraint: **k** ≥ 0.

On entry, **n** $=$ ⟨*value*⟩.
Constraint: **n** ≥ 1.

NE_INT_2

On entry, **n** $=$ ⟨*value*⟩ and **k** $=$ ⟨*value*⟩.
Constraint: **n** $\geq$ **k**.

NE_INTERNAL_ERROR

An internal error has occurred in this function. Check the function call and any array sizes. If the call is correct then please contact NAG for assistance.

An unexpected error has been triggered by this function. Please contact NAG.
See Section 2.7.6 in How to Use the NAG Library and its Documentation for further information.

NE_NO_LICENCE

Your licence key may have expired or may not have been installed correctly.
See Section 2.7.5 in How to Use the NAG Library and its Documentation for further information.

NE_REAL

On entry, **clevel** < 0.0 or **clevel** > 1.0: **clevel** $= \langle value \rangle$.

7 Accuracy

For most cases using the beta deviates the results should have a relative accuracy of $\max(0.5\mathrm{e}{-12}, 50.0 \times \epsilon)$ where ϵ is the ***machine precision*** (see nag_machine_precision (X02AJC)). 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 Parallelism and Performance

nag_binomial_ci (g07aac) is not threaded in any implementation.

9 Further Comments

None.

10 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.

10.1 Program Text

```
/* nag_binomial_ci (g07aac) Example Program.
 *
 * NAGPRODCODE Version.
 *
 * Copyright 2016 Numerical Algorithms Group.
 *
 * Mark 26, 2016.
 */

#include <stdio.h>
#include <nag.h>
#include <nag_stdlib.h>
#include <nagg07.h>

int main(void)
{

  /* Scalars */
  double clevel, phat, pl, pu;
  Integer exit_status = 0, k, n;
  NagError fail;

  INIT_FAIL(fail);

  printf("nag_binomial_ci (g07aac) Example Program Results\n");

  /* Skip heading in data file */
#ifdef _WIN32
  scanf_s("%*[^\n] ");
#else
  scanf("%*[^\n] ");
```

```
#endif

  printf("\n");
  printf(" Probability    Confidence Interval\n");
  printf("\n");
#ifdef _WIN32
  while (scanf_s("%" NAG_IFMT "%" NAG_IFMT "%lf%*[^\n] ", &n, &k, &clevel) !=
         EOF)
#else
  while (scanf("%" NAG_IFMT "%" NAG_IFMT "%lf%*[^\n] ", &n, &k, &clevel) !=
         EOF)
#endif
  {
    phat = (double) k / (double) n;
    /* nag_binomial_ci (g07aac).
     * Computes confidence interval for the parameter of a
     * binomial distribution
     */
    nag_binomial_ci(n, k, clevel, &pl, &pu, &fail);
    if (fail.code != NE_NOERROR) {
      printf("Error from nag_binomial_ci (g07aac).\n%s\n", fail.message);
      exit_status = 1;
      goto END;
    }

    printf("%10.4f      ( %6.4f , %6.4f )\n", phat, pl, pu);
  }

END:
  return exit_status;
}
```

10.2 Program Data

```
nag_binomial_ci (g07aac) Example Program Data
    1067     117    0.95           : n, k, clevel
     402      54    0.95
```

10.3 Program Results

```
nag_binomial_ci (g07aac) Example Program Results

 Probability    Confidence Interval

    0.1097      ( 0.0915 , 0.1300 )
    0.1343      ( 0.1025 , 0.1716 )
```