NAG C Library Function Document

nag_binomial_ci (g07aac)

1 Purpose

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

2 Specification

```c
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 parameter 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 $\alpha$ 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 $\beta_{a,b,c}$ is the deviate associated with the lower tail probability, $\delta$, of the beta distribution with parameters $a$ 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 parameters $n$ 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 $\chi^2$ 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 $\chi^2_{\nu,\delta}$ is the deviate associated with the lower tail probability, $\delta$, of the $\chi^2$ distribution with $\nu$ degrees of freedom.

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

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

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

where $\gamma_{\alpha,\beta,\delta}$ is the deviate associated with the lower tail probability, $\delta$, of the gamma distribution with shape parameter $\alpha$ and scale parameter $\beta$. These deviates are computed using \textit{nag\_deviates\_gamma\_dist (g01ffc)}.

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

The binomial variate with parameters $n$ 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 $z_\delta$ is the deviate associated with the lower tail probability, $\delta$, of the standard Normal distribution. These equations are solved using a quadratic equation solver.

### 4 References


Snedecor G W and Cochran W G (1967) \textit{Statistical Methods} Iowa State University Press

### 5 Parameters

1. $n$ – Integer

   On entry: the number of trials, $n$.

   Constraint: $n \geq 1$.

2. $k$ – Integer

   On entry: the number of successes, $k$.

   Constraint: $0 \leq k \leq n$.

3. clevel – double

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

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

4. $pl$ – double *

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

5. $pu$ – double *

   On exit: the upper limit, $p_u$, of the confidence interval.
6 Error Indicators and Warnings

NE_INT
On entry, \( k = \langle \text{value} \rangle \).
Constraint: \( k \geq 0 \).

On entry, \( n = \langle \text{value} \rangle \).
Constraint: \( n \geq 1 \).

NE_INT_2
On entry, \( n < k \): \( n = \langle \text{value} \rangle \) and \( k = \langle \text{value} \rangle \).

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

NE_REAL
On entry, \( \text{clevel} < 0.0 \) or \( \text{clevel} > 1.0 \): \( \text{clevel} = \langle \text{value} \rangle \).

NE_BAD_PARAM
On entry, parameter \( \langle \text{value} \rangle \) had an illegal value.

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 consult NAG for assistance.

7 Accuracy
For most cases using the beta deviates the results should have a relative accuracy of \( \max(0.5\varepsilon -12,50.0 \times \varepsilon) \) where \( \varepsilon \) 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 \( \text{clevel} \) 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
/* nag_binomial_ci (g07aac) Example Program.
* Copyright 2001 Numerical Algorithms Group.
*/

#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, k, n;
    NagError fail;

    INIT_FAIL(fail);
    exit_status = 0;
    Vprintf("g07aac Example Program Results\n");

    /* Skip heading in data file */
    Vscanf("%*\[\^
\] ");
    Vprintf("\n");
    Vprintf(" Probability Confidence Interval\n");
    Vprintf("\n");
    while ( scanf("%ld%ld%lf%*\[\^
\] ", &n, &k, &clevel) != EOF)
    {
        phat = (double) k / (double) n;
        g07aac(n, k, clevel, &pl, &pu, &fail);
        if (fail.code != NE_NOERROR)
        {
            Vprintf("Error from g07aac.
", fail.message);
            exit_status = 1;
            break;
        }
        Vprintf("%10.4f ( %6.4f , %6.4f )\n", phat, pl, pu);
    }
    return exit_status;
}

9.2 Program Data

<table>
<thead>
<tr>
<th>g07aac Example Program Data</th>
<th>n</th>
<th>k</th>
<th>clevel</th>
</tr>
</thead>
<tbody>
<tr>
<td>1067</td>
<td>117</td>
<td>0.95</td>
<td></td>
</tr>
<tr>
<td>402</td>
<td>54</td>
<td>0.95</td>
<td></td>
</tr>
</tbody>
</table>

9.3 Program Results

<table>
<thead>
<tr>
<th>g07aac Example Program Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Probability</td>
</tr>
<tr>
<td>0.1097</td>
</tr>
<tr>
<td>0.1343</td>
</tr>
</tbody>
</table>