NAG Fortran Library Routine Document

G07AAF

Note: before using this routine, please read the Users' Note for your implementation to check the interpretation of **bold italicised** terms and other implementation-dependent details.

1 Purpose

G07AAF computes a confidence interval for the parameter p (the probability of a success) of a binomial distribution.

2 Specification

```
SUBROUTINE G07AAF(N, K, CLEVEL, PL, PU, IFAIL)
INTEGER          N, K, IFAIL
real             CLEVEL, PL, PU
```

3 Description

Given the number of trials, n, and the number of successes, k, this routine 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 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 parameters a and b. These beta deviates are computed using G01FEF.

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 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 Hastings and Peacock (1975), page 70) 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 parameter α and scale parameter β. These deviates are computed using G01FFF.

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 Hastings and Peacock (1975), page 38.

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 C02AJF.

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

1: N – INTEGER *Input*

On entry: the number of trials, n.

Constraint: N $\geq$ 1.

2: K – INTEGER *Input*

On entry: the number of successes, k.

Constraint: $0 \leq$ K $\leq$ N.

3: CLEVEL – ***real*** *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 – ***real*** *Output*

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

5:    PU – ***real*** *Output*

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

6:    IFAIL – INTEGER *Input/Output*

On entry: IFAIL must be set to 0, −1 or 1. Users who are unfamiliar with this parameter should refer to Chapter P01 for details.

On exit: IFAIL = 0 unless the routine detects an error (see Section 6).

For environments where it might be inappropriate to halt program execution when an error is detected, the value −1 or 1 is recommended. If the output of error messages is undesirable, then the value 1 is recommended. Otherwise, for users not familiar with this parameter the recommended value is 0. **When the value −1 or 1 is used it is essential to test the value of IFAIL on exit.**

6   Error Indicators and Warnings

If on entry IFAIL = 0 or −1, explanatory error messages are output on the current error message unit (as defined by X04AAF).

Errors or warnings detected by the routine:

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.

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 X02AJF). 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

Note: the listing of the example program presented below uses *bold italicised* terms to denote precision-dependent details. Please read the Users' Note for your implementation to check the interpretation of these terms. As explained in the Essential Introduction to this manual, the results produced may not be identical for all implementations.

```
*      G07AAF Example Program Text
*      Mark 15 Release. NAG Copyright 1991.
*      .. Parameters ..
       INTEGER          NIN, NOUT
       PARAMETER        (NIN=5,NOUT=6)
*      .. Local Scalars ..
       real             CLEVEL, PHAT, PL, PU
       INTEGER          IFAIL, K, N
*      .. External Subroutines ..
       EXTERNAL         G07AAF
*      .. Intrinsic Functions ..
       INTRINSIC        real
*      .. Executable Statements ..
       WRITE (NOUT,*) 'G07AAF Example Program Results'
*      Skip heading in data file
       READ (NIN,*)
       WRITE (NOUT,*)
       WRITE (NOUT,*) ' Probability    Confidence Interval '
       WRITE (NOUT,*)
   20  READ (NIN,*,END=40) N, K, CLEVEL
       PHAT = real(K)/real(N)
       IFAIL = 0
*
       CALL G07AAF(N,K,CLEVEL,PL,PU,IFAIL)
*
       WRITE (NOUT,99999) PHAT, '( ', PL, ' , ', PU, ' )'
       GO TO 20
   40  STOP
*
99999 FORMAT (1X,F10.4,6X,A,F6.4,A,F6.4,A)
       END
```

9.2 Program Data

```
G07AAF Example Program Data
    1067     117    0.95           : N, K, CLEVEL
     402      54    0.95
```

9.3 Program Results

```
 G07AAF Example Program Results

  Probability    Confidence Interval

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