Non Asymptotic Binomial Confidence Intervals

David Harte*

20 May 2002

Abstract

Non asymptotic confidence intervals for a binomial proportion p can be calculated by the use of the F distribution, which is related to the binomial distribution via the incomplete beta function. These confidence intervals were tabulated and/or plotted by Pearson & Hartley (1962, Pages 33 and 77) and Neave (1978). Such confidence intervals can be easily calculated using modern statistical software. In this technical note, we derive the formulae for these confidence intervals.

Binomial Confidence Interval

Assume that n Bernoulli trials have been observed, and x of them resulted in "success". A sample estimate of the success probability is $\widehat{p} = x/n$.

Denote the $100(1-\alpha)\%$ confidence interval for p as (ϕ, ψ). It can be calculated as follows. The lower bound $\phi = 0$ if $x = 0$, and

$$\phi = \frac{x}{x + (n-x+1)F_{2(n-x+1),2x;1-\alpha/2}} \quad \text{if } x \neq 0,$$

where $F_{2(n-x+1),2x;1-\alpha/2}$ is the $100(1-\alpha/2)$ percentile of the F distribution with $2(n-x+1)$ numerator degrees of freedom and $2x$ denominator degrees of freedom. Similarly, $\psi = 1$ if $x = n$ and

$$\psi = \frac{(x+1)F_{2(x+1),2(n-x);1-\alpha/2}}{n-x+(x+1)F_{2(x+1),2(n-x);1-\alpha/2}} \quad \text{if } x \neq n.$$

Examples

1. Suppose $n = 20$ experiments were performed and $x = 1$ of them resulted in "success". Thus $\widehat{p} = 0.05$. The 95% non asymptotic confidence interval for the experimental success rate is calculated as follows:

$$\phi = \frac{1}{1 + 20F_{40,2;0.975}} = \frac{1}{1 + 20 \times 39.4729} = 0.0013$$

*Statistics Research Associates, PO Box 12 649, Wellington NZ

and

$$\psi = \frac{2F_{4,38;0.975}}{19 + 2F_{4,38;0.975}} = \frac{2 \times 3.1453}{19 + 2 \times 3.1453} = 0.2487,$$

i.e. $(0.0013, 0.2487)$. The normal approximation gives a 95% confidence interval as:

$$0.05 \pm 1.96\sqrt{\frac{(0.05)(0.95)}{20}} = 0.05 \pm (1.96)(0.0487) = (-0.0455, 0.1455)\ .$$

2. Now suppose that $n = 20$ and $x = 0$. Then $\phi = 0$, and

$$\psi = \frac{F_{2,40;0.975}}{20 + F_{2,40;0.975}} = \frac{4.0510}{20 + 4.0510} = 0.1684\ .$$

Derivation of the Solution

The values of ϕ and ψ satisfy the equations

$$\Pr\{X \geq x | X \sim \text{binomial}(n, \phi)\} = \alpha/2, \tag{1}$$

and

$$\Pr\{X \leq x | X \sim \text{binomial}(n, \psi)\} = \alpha/2, \tag{2}$$

respectively (see Agresti, 1990, Ex 3.29). In order to solve these equations for ϕ and ψ, we use relationships between the binomial distribution, the incomplete beta function and the F distribution.

Denote the incomplete beta function (see Abramowitz & Stegun, 1964, 26.5.2) by $I_x(a, b)$, where

$$\begin{aligned} I_x(a,b) &= \frac{1}{B(a,b)} \int_0^x t^{a-1}(1-t)^{b-1} dt \qquad 0 \leq x \leq 1 \\ &= 1 - I_{1-x}(b,a) \end{aligned}$$

and where

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)} \qquad a, b > 0.$$

Note also (Abramowitz & Stegun, 1964, 26.5.28) that

$$\Pr\{F_{\nu,\omega} \geq F\} = I_x(\omega/2, \nu/2) \quad \text{when} \quad x = \frac{\omega}{\omega + \nu F}\ . \tag{3}$$

The left hand side of Equation 1 can be rewritten (Abramowitz & Stegun, 1964, 26.5.24) as

$$\begin{aligned} \Pr\{X \geq x | X \sim \text{binomial}(n, \phi)\} &= \sum_{r=x}^{n} \binom{n}{r} \phi^r (1-\phi)^{n-r} \\ &= I_\phi(x, n-x+1) \qquad \text{if } x \neq 0\ . \end{aligned}$$

From Equation 3, $I_\phi(x, n-x+1) = \Pr\{F_{2(n-x+1),2x} \geq F\}$ when

$$\phi = \frac{x}{x + (n-x+1)F}\ .$$

We require $\Pr\{F_{2(n-x+1),2x} \geq F\} = \alpha/2$, thus $F = F_{2(n-x+1),2x;1-\alpha/2}$. If $x = 0$, then $\phi = 0$.

Similarly, the left hand side of Equation 2 can be rewritten as

$$\begin{aligned}
\Pr\{X \leq x | X \sim \text{binomial}(n, \psi)\} &= 1 - \sum_{r=x+1}^{n} \begin{pmatrix} n \\ r \end{pmatrix} \psi^r (1-\psi)^{n-r} \\
&= 1 - I_\psi(x+1, n-x) \qquad x \neq n \\
&= I_{1-\psi}(n-x, x+1).
\end{aligned}$$

Now $I_{1-\psi}(n-x, x+1) = \Pr\{F_{2(x+1),2(n-x)} \geq F\}$ when

$$1 - \psi = \frac{2(n-x)}{2(n-x) + 2(x+1)F} \ .$$

That is, when

$$\psi = \frac{(x+1)F}{n-x+(x+1)F} \ .$$

We require F such that $\Pr\{F_{2(x+1),2(n-x)} \geq F\} = \alpha/2$, thus $F = F_{2(x+1),2(n-x);1-\alpha/2}$. If $x = n$, then $\psi = 1$.

References

Abramowitz, M. & Stegun, I.A. (1964). *Handbook of Mathematical Functions.* Dover, New York.

Agresti, A. (1990). *Categorical Data Analysis.* Wiley, New York.

Neave, H.R. (1978). *Statistics Tables for Mathematicians, Engineers, Economists and the Behavioural and Management Sciences.* George Allen & Unwin, London.

Pearson, E.S. & Hartley, H.O. (1962). *Biometrika Tables for Statisticians.* Cambridge University Press, Cambridge.