Copper-Pearson Interval

Scotia Bank Group

August 18, 2016

The Clopper-Pearson interval is a method for calculating binomial confidence intervals.

CP is an exact method, because it is based on the cumulative probabilities of the binomial distribution.

However, the intervals are not exact in the sense that the discontinuous nature of the binomial distribution precludes any interval with exact coverage for all population proportions.

This interval never has less than the nominal coverage for any population proportion, but that means that it is usually conservative.

For example, the true coverage rate of a 0.95 Clopper-Pearson interval may be well above 0.95, depending on n and θ. Thus the interval may be wider than it needs to be to achieve 0.95 confidence.

In contrast, it is worth noting that other confidence bounds may be narrower than their nominal confidence width, i.e., the Normal Approximation (or "Standard") Interval, Wilson Interval, Agresti-Coull Interval, etc., with a nominal coverage of 0.95 may in fact cover less than 0.95.

The Clopper-Pearson interval takes the form:

$$S_{\leq} \cap S_{\geq} \quad \text{or equivalently} \quad (\inf S_{\geq}\,,\, \sup S_{\leq})$$

with

$$S_{\leq} := \left\{\theta \middle| P\left[\mathrm{Bin}\left(n;\theta\right) \leq X\right] > \frac{\alpha}{2}\right\} \quad \text{and} \quad S_{\geq} := \left\{\theta \middle| P\left[\mathrm{Bin}\left(n;\theta\right) \geq X\right] > \frac{\alpha}{2}\right\},$$

where $0 \leq X \leq n$ is the number of successes observed in the sample and $Bin(n;\theta)$ is a binomial random variable with n trials and probability of success θ.

Because of a relationship between the cumulative binomial distribution and the beta distribution, the Clopper-Pearson interval is sometimes presented in an alternate format that uses quantiles from the beta distribution.

$$B\left(\frac{\alpha}{2}; x, n-x+1\right) < \theta < B\left(1-\frac{\alpha}{2}; x+1, n-x\right)$$

where x is the number of successes, n is the number of trials, and $B(p;v,w)$ is the $p-$th quantile from a beta distribution with shape parameters v and w.

When x is either 0 or n, closed-form expressions for the interval bounds are available: when $x=0$ the interval is

$$I_n := \left(0, 1-\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}\right)$$

and when $x=n$ it is

$$I_n := \left(\left(\frac{\alpha}{2}\right)^{\frac{1}{n}}, 1\right).$$

The beta distribution is, in turn, related to the F-distribution so a third formulation of the Clopper-Pearson interval can be written using F quantiles:

$$\left(1+\frac{n-x}{[x+1]\ F\big[\frac{\alpha}{2};2(x+1),2(n-x)\big]}\right)^{-1}<\theta<\left(1+\frac{n-x+1}{x\ F\big[1-\frac{1}{2}\alpha;2x,2(n-x+1)\big]}\right)^{-1}$$

where x is the number of successes, n is the number of trials, and $F(c;d_1,d_2)$ is the $1-c$ quantile from an F-distribution with d_1 and d_2 degrees of freedom.