**Section 1.1.2 – Inference for the probability of success (continued)**

Wald confidence interval

Because  has an approximate normal distribution (mean: θ, variance: ) for a large sample, we can rewrite this as a standardized statistic:

After recording video:  means “approximate distribution” and N(0,1) is shorthand notation for a normal distribution with a mean of 0 and variance of 1 (i.e., a standard normal). Both of these uses of symbols are common in statistics but some students may not have seen them before.



Also, because we have a probability distribution here, we can quantify with a level of certainty that observed values of the statistic are within a particular range:



where  is the 1 – α/2 quantile from a standard normal. For example, if α = 0.05, we have Z0.975 = 1.96:

> qnorm(p = 1-0.05/2, mean = 0, sd = 1)

[1] 1.959964

Notes:

* Be careful with the subscript for Z. The value in the subscript is the area to the LEFT of the quantile. Some introductory statistics textbooks will use the subscript to denote the area to the RIGHT of the quantile.
* I specially chose  and  for symmetry. Of course, .

If we rearrange items within the P(⋅), we obtain



Thus, if α is chosen to be small, we are fairly certain the expression within P(⋅) will hold true. When we substitute the observed values of  and  into the expression, we obtain the (1 – α)100% Wald confidence interval for θ as



Wald (1943) developed this result. Notice this interval follows the typical form of a confidence interval for a parameter:

Estimator ± (distributional value)\*(standard deviation of estimator)

Confidence interval for π

Because  is a maximum likelihood estimator, we can use a Wald confidence interval for π:



Equivalently, we can write the interval as



When  is close to 0 or 1, two problems may occur:

1. Calculated limits may be less than 0 or greater than 1, which is outside the boundaries for a probability.
2. When  = 0 or 1,  = 0 for n > 0. This leads to the lower and upper limits to be exactly the same (0 for  = 0 or 1 for  = 1).

Example: Field goal kicking (CIpi.R)

Suppose  = w = 4 and n = 10. The 95% confidence interval is



0.0964 < π < 0.7036

R code and output:

> w <- 4

> n <- 10

> alpha <- 0.05

> pi.hat <- w/n

> var.wald <- pi.hat\*(1-pi.hat)/n

> lower <- pi.hat - qnorm(p = 1-alpha/2, mean = 0, sd = 1) \* sqrt(var.wald)

> upper <- pi.hat + qnorm(p = 1-alpha/2, mean = 0, sd = 1) \* sqrt(var.wald)

> round(data.frame(lower, upper), digits = 4)

 lower upper

1 0.0964 0.7036

> #Quicker

> round(pi.hat + qnorm(p = c(alpha/2, 1-alpha/2)) \*

 sqrt(var.wald), 4)

[1] 0.0964 0.7036

A set of R code will be introduced later that makes the computational process a little easier.

Interpretations:

* With 95% confidence, true probability of success is between 0.0964 and 0.7036.
* The 95% confidence interval is 0.0964 < π < 0.7036. We would expect that 95% of all similarly constructed intervals to contain π.

Incorrect interpretations:

* The probability that π is between 0.0964 and 0.7036 is 95%.
* 95% of the time π is between 0.0964 and 0.7036.
* The 95% confidence interval is 0.0964 < π < 0.7036. We would expect that 95% of all π values to be within these limits.

What’s the difference between “With 95% confidence” and “The probability … is 95%”?

Problems!!!

1. Remember the interval “works” if the sample size is large. The field goal kicking example has n = 10 only!
2. The discreteness of the binomial distribution often makes the normal approximation work poorly even with large samples.

The result is a confidence interval that is often too “liberal”. This means when 95% is stated as the confidence level, the true confidence level is often lower.

On the other hand, there are “conservative” intervals. These intervals have a true confidence level larger than the stated level.

The problems with this particular confidence interval have been discussed for a long time in the statistical literature. There have been many, many alternative confidence intervals for π proposed. The reason is because there is not just ONE interval that performs best ALL of the time. While research continues in this area, I think Brown et al. (*Statistical Science*, 2001) remains as the most comprehensive paper.

Below are my recommendations:

* Wilson interval (Wilson, 1927)



with .

Where does interval come from?

Consider the hypothesis test for H0:π = π0 vs. Ha:π ≠ π0 using the test statistic of



The limits of the Wilson confidence interval come from “inverting” the test. This means finding the set of π0 such that



is satisfied. Through solving a quadratic equation, the interval limits are derived.

* Agresti-Coull interval (Agresti and Coull, 1998)

The confidence interval is



This is essentially a Wald interval where we add  successes and  failures to the observed data. In fact, when α = 0.05, Z1-α/2 = 1.96 ≈ 2. Then

.

Thus, two successes and two failures are added. Also, notice how



can be thought of as an adjusted estimate of π. Notice how this is used in the Wilson interval. For values of w less than 0.5, . For values of w greater than 0.5, . Think about how this affects the interval!

* Clopper-Pearson interval – This will be introduced after we examine “true confidence level”.

For VERY extreme values of π, the Wilson interval can be liberal and the Agresti-Coull interval can be conservative. However, there are small fixes that can be performed to improve their performance. Please see the book for a discussion (these fixes are usually not implemented by R functions).

Example: Field goal kicking (CIpi.R)

Below is the code used to calculate each confidence interval.

> pi.tilde <- (w + qnorm(p = 1-alpha/2)^2 / 2) / (n + qnorm(p = 1-alpha/2)^2)

> pi.tilde

[1] 0.4277533

> #Wilson C.I.

> round(p.tilde + qnorm(p = c(alpha/2, 1-alpha/2)) \*

 sqrt(n) / (n+qnorm(p = 1-alpha/2)^2) \* sqrt(pi.hat\*(1-

 pi.hat) + qnorm(1-alpha/2)^2/(4\*n)), 4)

[1] 0.1682 0.6873

> #Agresti-Coull C.I.

> var.ac <- p.tilde\*(1-p.tilde) / (n+qnorm(p = 1-alpha/2)^2)

> round(p.tilde + qnorm(p = c(alpha/2, 1-alpha/2)) \*

 sqrt(var.ac), 4)

[1] 0.1671 0.6884

How useful would these confidence intervals be actual application?

The binom package in R provides a simple function to do these calculations as well. Here is an example of how to use the function:

> library(package = binom)

> binom.confint(x = w, n = n, conf.level = 1 – alpha, methods = "all")

 method x n mean lower upper

1 agresti-coull 4 10 0.4000000 0.16711063 0.6883959

2 asymptotic 4 10 0.4000000 0.09636369 0.7036363

3 bayes 4 10 0.4090909 0.14256735 0.6838697

4 cloglog 4 10 0.4000000 0.12269317 0.6702046

5 exact 4 10 0.4000000 0.12155226 0.7376219

6 logit 4 10 0.4000000 0.15834201 0.7025951

7 probit 4 10 0.4000000 0.14933907 0.7028372

8 profile 4 10 0.4000000 0.14570633 0.6999845

9 lrt 4 10 0.4000000 0.14564246 0.7000216

10 prop.test 4 10 0.4000000 0.13693056 0.7263303

11 wilson 4 10 0.4000000 0.16818033 0.6873262