**Likelihood ratio tests (LRTs)**

The LRT is a general way to test hypotheses. The LRT statistic, Λ, is the ratio of two likelihood functions. The numerator is the likelihood function maximized over the parameter space restricted under the null hypothesis. The denominator is the likelihood function maximized over the combined parameter space for the null and alternative hypotheses. The test statistic is written as:



Wilks (1935, 1938) showed that –2log(Λ) can be approximated by a chi-square distribution with u degrees of freedom for a large sample and under Ho, where u is the difference in dimension between the alternative and null hypothesis parameter spaces. The Λ statistic here is a random variable because it is a function of Y1, …, Yn.

Questions:

* Suppose Λ is close to 1, what does this say about Ho? Explain.
* Suppose Λ is close to 0, what does this say about Ho? Explain.
* When using –2log(Λ), will large or small value values indicate Ho is false?
* Is this a two-tail, left-tail, or right-tail test?

Example: Field goal kicking (LikelihoodFunction.R)

Continuing the field goal example, suppose the hypothesis test

Ho: π = 0.5

Ha: π ≠ 0.5

is of interest. Remember that = w = 4 and n = 10.

The numerator of Λ is the maximum possible value of the likelihood function under the null hypothesis. Because π = 0.5 is the null hypothesis, the maximum can be found by substituting π = 0.5 into the likelihood function:



Then



The denominator of Λ is the maximum possible value of the likelihood function under the null OR alternative hypotheses. Because this includes all possible values of π here, the maximum is achieved when the MLE is substituted for π in the likelihood function! As shown previously, the maximum value is 0.001194.

Therefore,



NOTE: The observed value of the LRT statistic is denoted by an uppercase letter rather than a lowercase letter. This is the standard way it is used. As you take more statistics courses, you will see that uppercase and lowercase notation is not emphasized because an individual has enough experience with statistics that it should be clear if the random variable or the observed value is being used.

–2log(Λ) = –2log(0.8179) = 0.4020.

The critical value is  = 3.84 using α = 0.05. The p-value is 0.53.

> qchisq(p = 0.95, df = 1)

[1] 3.841459

> 1 - pchisq(q = -2\*log(Lambda), df = 1)

[1] 0.5256929

Steps:

1. H0: π = 0.5 vs. Ha: π ≠ 0.5
2. -2log(Λ) = 0.4020
3.  = 3.84
4.



Because 0.4020 < 3.84, do not reject Ho.

1. There is not sufficient evidence to reject the hypothesis that the probability of success is equal to 0.5.

What happens if there is more than one parameter of interest?

Suppose θ1 and θ2 are the parameters and we observe y1, …, yn. If the hypotheses are

Ho: θ1 = θ10 and θ2 = θ20

Ha: Not all equal in H0

the LRT test statistic is



where  and  are the MLEs. If the hypotheses were instead

Ho: θ1 = θ10

Ha: θ1 ≠ θ10

so that θ2 is not specified, the numerator for Λ changes. The statistic becomes



where  is the MLE assuming θ1 = θ10.

Likelihood ratio CIs

Many hypothesis tests can be “inverted” to produce CIs. We will focus on the LRT. To illustrate the process, consider a situation with only one parameter θ. If we were to perform the hypothesis test of

Ho: θ = θ0

Ha: θ ≠ θ0

for some particular numerical value θ0, we would use

,

where  is the MLE. Values of –2log(Λ) ≤  indicate there is not sufficient evidence to reject H0. Values of
–2log(Λ) >  give sufficient evidence to reject H0.

The LR CI is all possible values of θ0 such that

–2log(Λ) ≤ 

is satisfied. Due to the form of likelihood functions, these all possible values will be one interval, like 2 < θ < 3, rather than a set of disjoint values/intervals.

For most parameters and distributions of interest, iterative numerical methods need to be used to find the interval.

Example: Field goal kicking (LikelihoodFunction.R)

Continuing the field goal example, we would like a CI for π. The LRT statistic is



Working with the transformation, we have



The LR CI is all possible values of π0 such that

–2log(Λ) ≤ 

Unfortunately, a nice closed-form expression does not exist for the set of π0 values. Therefore, we need to use numerical iterative methods to find the interval.

> find.root <- function(pi.0, w, n) {

 pi.hat <- w/n

 -2 \* (w\*log(pi.0/pi.hat) + (n - w)\*log((1 - pi.0)/(1 –

 pi.hat))) - qchisq(p = 0.95, df = 1)

 }

> # Test function

> find.root(pi.0 = 0.2, w = w, n = n)

[1] -1.748466

> curve(expr = find.root(pi.0 = x, w = w, n = n), xlim =

 c(0,1), col = "red", xlab = "pi", ylab = "-2log(Lambda)

 - chi-square\_1,0.05")

> abline(h = 0, col = "blue", lty = "dotted")



> # Find interval

> lower <- uniroot(f = find.root, interval = c(0.01, w/n),

 w = w, n = n)

> lower$root

[1] 0.1456459

> upper <- uniroot(f = find.root, interval = c(w/n, 0.99),

 w = w, n = n)

> upper$root

[1] 0.7000256

The 95% LR CI is 0.15 < π < 0.70.

Alternatively, one could perform a simple search as follows.

> # Calculate -2log(Lambda)

> TranLambda <- function(pi.0, w, n) {

 pi.hat <- w/n

 -2 \* (w \*log(pi.0/pi.hat) + (n - w)\*log((1 - pi.0)/(1 –

 pi.hat)))

 }

> # Test function

> TranLambda(pi.0 = 0.2, w = w, n = n)

[1] 2.092993

> # Calculate -2log(Lambda) for a set of pi potential

 values

> pi <- seq(from = 0.01, to = 0.99, by = 0.01)

> save.value <- numeric(length(pi))

> counter <- 1

> for(pi.0 in pi) {

 save.value[counter] <- TranLambda(pi.0 = pi.0, w = w, n

 = n)

 counter <- counter + 1

 }

> # Check if -2log(Lambda) <= chisquare(0.95, 1)

> save.check <- save.value <= qchisq(p = 0.95, df = 1)

> lower <- min(pi[save.check])

> upper <- max(pi[save.check])

> data.frame(lower, upper)

 lower upper

1 0.15 0.7

Lastly, one could use the binom.confint() function from the binom package!

> library(binom)

> binom.confint(x = w, n = n, conf.level = 0.95, methods =

 "lrt")

 method x n mean lower upper

1 lrt 4 10 0.4 0.1456425 0.7000216

Side note: There are many other CI methods for π.

> binom.confint(x = w, n = n, conf.level = 0.95, 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

For example, the “asymptotic” method is the Wald CI discussed earlier in this section. Notice how different the lower bound is for this interval compared to the LR CI. In general, LR CIs are better than Wald CIs for this and other parameters of interest.

What does “better” mean?

What happens when there is more than one parameter in the likelihood function?

We calculate what is known as profile LR CI. This follows a similar procedure as we saw before when θ1 and θ2 were the parameters, and we wanted to perform a test only for θ1. For the interval, we now find the set of all θ10 numerical values such that ≤  and

.

where  is the MLE assuming θ1 = θ10.