Empirical Power Study

Matthew Maslow

5/10/2023

# Introduction

In my empirical study, I will be taking a looking at several scenarios for a binomial test for proportion. More specifically, also written in the hypothesis below, for the proportion, p = 0.65, where I will be testing different alternate hypotheses, to compare their powers. Where power is the probability of rejecting the null hypothesis, and we are going to be using a α=0.05 significance level. Furthermore, I will be constructing a test for the binomial distribution, then I will construct a test where I will convert the distribution to the standard normal. For the tests, I must keep in account that a binomial sample is a discrete distribution, therefore I must implement a continuity correction with my true proportions/alternative hypotheses. With the standard normal distribution of the binomial samples, since it is continuous, we do not have to worry about a continuity correction. Lastly, throughout each of the cases, I will be running both test under different sample sizes (small, medium, large, and very-large) to compare the outputs of each scenario, as well comparing the outcomes of the binomial versus standard normal distributions.

# Hypothesis

library(tidyverse)

Binomial test of a proportion with….

Test for less than so would be left-tail test for different “true” alternatives.

# Case 1: Small Sample Size (n = 5)

null = 0.65 # null hypothesis  
true\_props = seq(0.1, null + 0.1, by = 0.05) # true prop’s/alt hypotheses   
nsim = 1000 # specify number of random data-sets to be generated for each case (to approximate the power)  
n1 = 5 # Sample size  
case1\_power = rep(0, length(true\_props) ) # sets up place to store powers   
case1\_NormPower = rep(0, length(true\_props) )  
# loop through all of the possible alternatives  
for(i in 1:length(true\_props)){   
 rejects = 0 # sets up place to store # of null.hyp rejections   
 rejectsNorm = 0  
 for (j in 1:nsim){# loop through the nsim datasets you will generate  
 # generate one random dataset for your scenario  
 x1 = rbinom(1, n1, true\_props[i])   
 # update your counter if you reject Ho at 5% level  
 if (binom.test(x1, n1, p = null, alternative = "less")$p.value < 0.05) rejects = rejects + 1

# normalizing test  
 phat = x1/n1 # calculations for p-hat  
 z = (phat-null)/sqrt((null\*(1-null))/n1) # test statistic

# update your counter if you reject Ho at 5% level  
 if(pnorm(z, 0, 1) < 0.05) rejectsNorm = rejectsNorm + 1 }  
 case1\_power[i] = rejects/nsim # determine the proportion of times you rejected Ho out of nsim replications  
 case1\_NormPower[i] = rejectsNorm/nsim }

# Case 2: Medium Sample Size (n = 50)

null = 0.65  
true\_props = seq(0.1, null + 0.1, by = 0.05)  
nsim = 1000   
n2 = 50   
case2\_power = rep(0, length(true\_props) )   
case2\_NormPower = rep(0, length(true\_props) )   
for (i in 1:length(true\_props)) {  
 rejects = 0   
 rejectsNorm = 0  
 for (j in 1:nsim) {  
 x2 = rbinom(1, n2, true\_props[i])   
 if (binom.test(x2, n2, p = null, alternative = "less")$p.value < 0.05) rejects = rejects + 1  
 phat = x2/n2  
 z = (phat-null)/sqrt((null\*(1-null))/n2)  
 if(pnorm(z, 0, 1) < 0.05) rejectsNorm = rejectsNorm + 1 }  
 case2\_power[i] = rejects/nsim  
 case2\_NormPower[i] = rejectsNorm/nsim }

# Case 3: Large Sample Size (n = 100)

null = 0.65   
true\_props = seq(0.1, null + 0.1, by = 0.05)  
nsim = 1000  
n3 = 100  
case3\_power = rep(0, length(true\_props) )   
case3\_NormPower = rep(0, length(true\_props) )  
for (i in 1:length(true\_props)) {  
 rejects = 0   
 rejectsNorm = 0  
 for (j in 1:nsim) {  
 x3 = rbinom(1, n3, true\_props[i])   
 if (binom.test(x3, n3, p = null, alternative = "less")$p.value < 0.05) rejects = rejects + 1  
 phat = x3/n3  
 z = (phat-null)/sqrt((null\*(1-null))/n3)  
 if(pnorm(z, 0, 1) < 0.05) rejectsNorm = rejectsNorm + 1  
 }  
 case3\_power[i] = rejects/nsim   
 case3\_NormPower[i] = rejectsNorm/nsim  
}

# Case 4: Very Large Sample Size (n = 1000)

null = 0.65   
true\_props = seq(0.1, null + 0.1, by = 0.05)  
nsim = 1000   
n4 = 1000   
case4\_power = rep(0, length(true\_props) )  
case4\_NormPower = rep(0, length(true\_props) )  
  
for (i in 1:length(true\_props)) {  
 rejects = 0  
 rejectsNorm = 0  
 for (j in 1:nsim) {  
 x4 = rbinom(1, n4, true\_props[i])   
 if (binom.test(x4, n4, p = null, alternative = "less")$p.value < 0.05) rejects = rejects + 1  
 phat = x4/n4  
 z = (phat-null)/sqrt((null\*(1-null))/n4)  
 if(pnorm(z, 0, 1) < 0.05) rejectsNorm = rejectsNorm + 1  
 }  
 case4\_power[i] = rejects/nsim  
 case4\_NormPower[i] = rejectsNorm/nsim }

# Data Set of Powers

results\_dataframe = data.frame(true\_props = true\_props,  
 case1\_power = case1\_power, case1\_NormPower = case1\_NormPower,  
 case2\_power = case2\_power, case2\_NormPower = case2\_NormPower,  
 case3\_power = case3\_power,case3\_NormPower = case3\_NormPower,  
 case4\_power = case4\_power, case4\_NormPower = case4\_NormPower  
 )

binom <- results\_dataframe %>%  
 select(true\_props, case1\_power, case2\_power, case3\_power, case4\_power)  
pander::pander(binom)

Binomial distribution…

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| true\_props | case1\_power(5) | case2\_power(50) |  | case3\_power(100) | case4\_power(1000) |
| 0.10 | 0.596 | 1.000 |  | 1.000 | 1.000 |
| 0.15 | 0.426 | 1.000 |  | 1.000 | 1.000 |
| 0.20 | 0.347 | 1.000 |  | 1.000 | 1.000 |
| 0.25 | 0.269 | 1.000 |  | 1.000 | 1.000 |
| 0.30 | 0.164 | 1.000 |  | 1.000 | 1.000 |
| 0.35 | 0.113 | 0.995 |  | 1.000 | 1.000 |
| 0.40 | 0.072 | 0.972 |  | 1.000 | 1.000 |
| 0.45 | 0.047 | 0.881 |  | 0.992 | 1.000 |
| 0.50 | 0.033 | 0.669 |  | 0.903 | 1.000 |
| 0.55 | 0.016 | 0.380 |  | 0.649 | 1.000 |
| 0.60 | 0.011 | 0.155 |  | 0.238 | 0.938 |
| 0.65 | 0.003 | 0.038 |  | 0.033 | 0.062 |
| 0.70 | 0.001 | 0.006 |  | 0.004 | 0.000 |
| 0.75 | 0.002 | 0.000 |  | 0.000 | 0.000 |
|  |  |  |  |  |  |
|  |  |  |  |  |  |

norm <- results\_dataframe %>%  
 select(true\_props, case1\_Norm, case2\_Norm, case3\_Norm, case4\_Norm)  
pander::pander(norm)

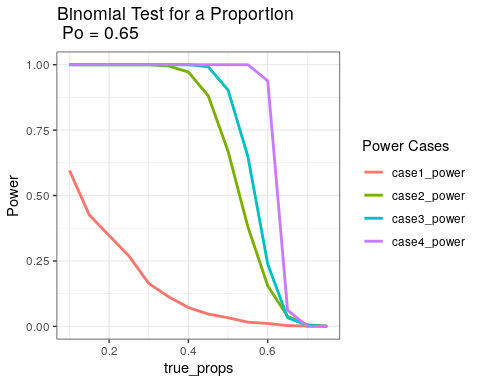
Standardized Binomial distribution…

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| true\_props | case1\_Norm(5) | case2\_Norm(50) |  | case3\_Norm(100) | case4\_Norm(1000) |
| 0.10 | 0.908 | 1.000 |  | 1.000 | 1.000 |
| 0.15 | 0.849 | 1.000 |  | 1.000 | 1.000 |
| 0.20 | 0.733 | 1.000 |  | 1.000 | 1.000 |
| 0.25 | 0.628 | 1.000 |  | 1.000 | 1.000 |
| 0.30 | 0.547 | 1.000 |  | 1.000 | 1.000 |
| 0.35 | 0.435 | 0.998 |  | 1.000 | 1.000 |
| 0.40 | 0.331 | 0.961 |  | 1.000 | 1.000 |
| 0.45 | 0.225 | 0.879 |  | 0.991 | 1.000 |
| 0.50 | 0.188 | 0.669 |  | 0.934 | 1.000 |
| 0.55 | 0.117 | 0.407 |  | 0.683 | 1.000 |
| 0.60 | 0.108 | 0.175 |  | 0.321 | 0.949 |
| 0.65 | 0.046 | 0.042 |  | 0.052 | 0.054 |
| 0.70 | 0.026 | 0.003 |  | 0.000 | 0.000 |
| 0.75 | 0.017 | 0.000 |  | 0.000 | 0.000 |
|  |  |  |  |  |  |

# Comparing Plots

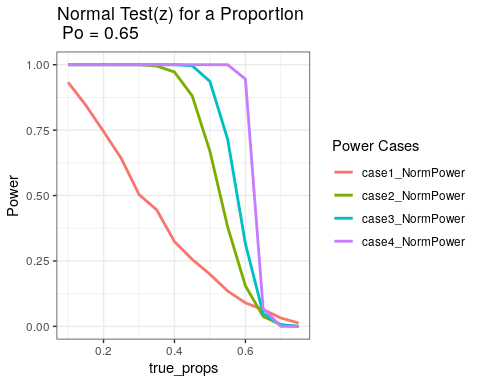
results\_dataframe %>%  
 pivot\_longer(., c(case1\_power, case2\_power, case3\_power, case4\_power),  
 names\_to = "Power Cases", values\_to = "Power") %>%  
 ggplot(., aes(x = true\_props, y = Power, col = `Power Cases`) ) +  
 geom\_line(lwd = 1) + labs(title = "Binomial Test for a Proportion\n Po = 0.65") + theme\_bw()

# Single Proportion Test for Binomial Distribution (Po = 0.65)



# Normal  
results\_dataframe %>%  
 pivot\_longer(., c(case1\_NormPower, case2\_NormPower, case3\_NormPower, case4\_NormPower),  
 names\_to = "Power Cases", values\_to = "Power") %>%  
 ggplot(., aes(x = true\_props, y = Power, col = `Power Cases`) ) +  
 geom\_line(lwd = 1) + labs(title = "Normal Test(z) for a Proportion\n Po = 0.65") + theme\_bw()

# Single Proportion Test for Standardized Binomial Distribution (Po = 0.65)



# Conclusion:

In this empirical study, I considered numerous approaches to fix any unmet condition or issues with continuous versus discrete distribution was the Agresti interval, the standard normal model(standardizing), or testing with different sample sizes. In all the scenarios, the test is implemented in a way to include a continuity correction as the binomial distribution is discrete and will help approximate the powers.

Looking at binomial versus the standard normal test, we can notice a few similarities as well as some differences. More specifically, if we look at the results of the binomial test, we will notice a few powers that do not make sense in the context of the problem, which poses as type I or II errors. Therefore, we standardize the binomial distribution converting it from discrete to continuous distribution to keep a more consistent trend/pattern in data. As a result, we can get more accurate results that make sense within the context of the problem, while satisfying any unmet conditions as well.

For the four different cases, I tested different sample sizes, small, medium, large, and very large. For each sample size, I ran a binomial and standard normal test. All the cases met their conditions except the small sample size of n=5, as they're less than ten fails and successes. Therefore, standardizing this case was helpful as it corrected for the inconsistency in the data allowing me to get more accurate results for the smaller sample size, making more sense with the true proportions. Furthermore, the general pattern we are seeing for each case is the power is starting out high, closer to one, then as the true proportion increases getting closer to the hypothesized null value, they're getting closer and closer to a power of 0.05 until we get to true proportions above the null and their powers become significant. This is the case as, we are testing for values in the range of 0.1 to 0.75 (sequence by 0.05), with the expression (Po=0.65) < (true proportions). So, for the proportions 0.1 through 0.6 should be insignificant powers as they are not greater than Po=0.65. Another thing that can be noticed as we increase the sample size, the powers of the true proportions that are considered insignificant start to get closer. This can be seen starting in the second, third, and fourth cases, where the number of true proportions starting at 0.1 and then eventually up to 0.55 all have powers equal to 1.000, which is very insignificant, and the number of their powers increases as the sample size increases. It seems that as you increase the sample size for a single proportion hypothesis test, with define a clear line between right and wrong, meaning that it makes it more probable that we are to reject the null for the true alternative values/true proportions.

Interestingly enough if you look at the true proportion of p = 0.65, the hypothesized null value, the binomial distribution test infers that testing the null against the null will prove significant, however, this does not make sense since 0.65 is not less than 0.65. This is the reason why we implemented the continuity correction, as binomial is a discrete distribution, therefore is inclusive meaning that when using a less than sign with discrete it means less than or equal. Therefore, this is the reason why the power for p = 0.65, which is equal to the null, is significant because of the rules of discrete distributions. This is why we added the continuity correction to account for this confusion. A similar thing can be seen for the standard normal test at the true proportion, p = 0.65, as we are only standardizing the binomial distribution to convert it to a continuous distribution. However, doing this does not correct it entirely, it only makes the power a little less significant at that proportion. Also, looking at the plots the larger sample sizes have much greater densities over 1.000. Also on the plots, binomial versus normal, we cannot see much of a difference between them other than with the smaller sample size.

To wrap things up, in general standardizing the binomial distribution helps with the consistency and logicality of the results, as binomial is discrete, and corrects for any continuity errors with inclusive versus exclusive with the discrete and continuous distributions. Furthermore, increasing sample size seems to help with accuracy and defines which points are clearly significant versus insignificant. In conclusion, I think it is best to have the binomial, distribution standardized to fix the consistency, as well as having a sample that is large enough to satisfy the conditions, to avoid skewness in the results.