Homework 2

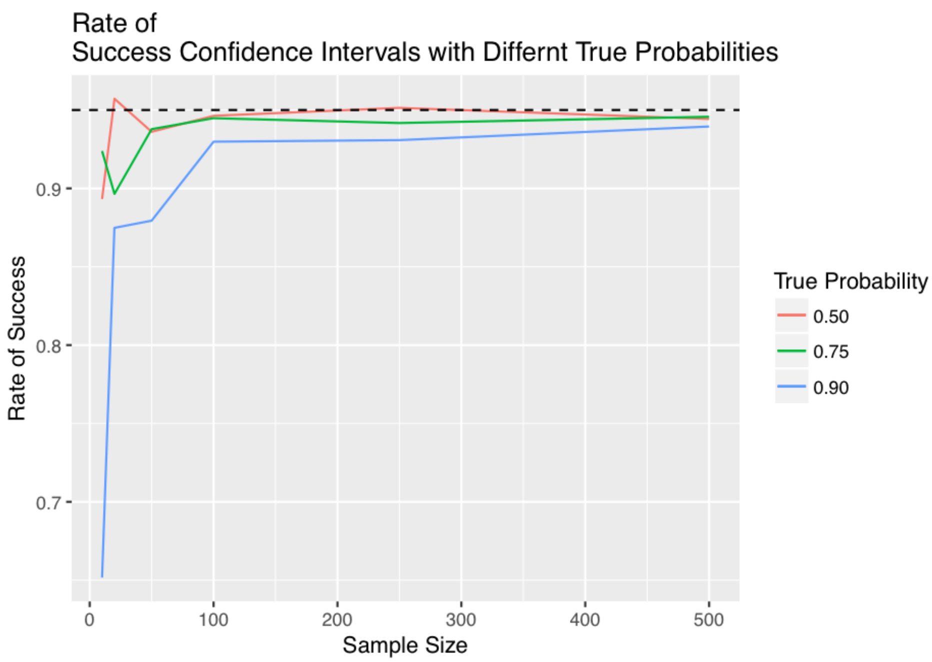
Tyler Poelking

1/29/2017

R Markdown

```r
#95% C.I z stat
zstat=qnorm(1-0.05/2)
n=10000
grid = data.frame()


#vector to store rows
row=vector(length=3)

for(m in c(10,20,50,100,250,500)){
    iter = 1
    for(p in c(0.5,0.75,0.9)){


#generate random binomial variables
v=rbinom(n=n,size=m, prob=p)

#calc sample proportions
sp=v/m

#CI from data
se=sqrt((sp*(1-sp))/m)

#Vector of lower bounds
lb=sp-(zstat*se)

#Vector of upper bounds
ub=sp+(zstat*se)

#Store cases where true p outside of generated CI
data=vector(length=n)
for(i in 1:n){
    if(lb[i]<p && ub[i]>p){
        data[i] = 1
    }
}
#calculate proportion true p inside generated CI
result = sum(data)/n

row[iter] = result
iter=iter+1
    }
    grid = rbind(grid, row)
}
```

```r
rownames(grid) = c("10","20","50","100","250","500")
colnames(grid) = c("0.5","0.75","0.9")

library(ggplot2)
plot = ggplot() +
geom_line(data = grid, aes(x = c(10,20,50,100,250,500), y = grid$`0.5`, color = "0.50")) +
geom_line(data = grid, aes(x = c(10,20,50,100,250,500), y = grid$`0.75`, color = "0.75"))+
geom_line(data = grid, aes(x = c(10,20,50,100,250,500), y = grid$`0.9`, color = "0.90"))+
geom_hline(aes(yintercept= 0.95),linetype = 2,
colour= 'black', show.legend =FALSE) +
labs(color="True Probability", title="Rate of
Success Confidence Intervals with Differnt True Probabilities") +
xlab('Sample Size') +
ylab('Rate of Success')
plot
```



From the data and the plot, one can observe that as the sample size increases the success rate of the confidence intervals tends to approach 0.95. When the true probability is lower, the rate of success of the confidence intervals starts out highest and stays as such. In other words, the 0.5 true probability lines tends to stay above the 0.75 line for all sample sizes. This is also true for the 0.75 true probability line staying above the 0.90 probability line for all sample sizes.