

ECON 6910
Applied Econometrics
Philip Shaw
Problem Set 6
Due Date: April 8, 2019

Problem 1

Consider the cumulative distribution function (CDF) estimator discussed in class:

$$F_n(x) = n^{-1} \sum_{i=1}^n I(X_i \leq x) \quad (1)$$

where $F_n(x)$ is an estimator for the population CDF $F(x)$.

a. Write a R-script file called `unicdf.R` that estimates $F(x)$ using equation 1. The code should take as inputs the data X_i for $i = 1, \dots, n$ and the point x for which the CDF is to be estimated. It should return the estimated value for the CDF at the point x .

b. Now draw data in R from a standard normal distribution using the $x = rnorm(n)$ command for a sample size $n = 50$. Estimate $F(0)$ for your draw of data. What is your estimated value $F_n(0)$? Compare this to the truth. How close are you?

c. Now repeat your experiment above for $trials = 1000$ with the help of a for loop in R. For loops can be constructed as follows:

```
for (i in 1:trials){  
  x=rnorm(n)  
  cdfs[i]=unicdf(x,0)  
}
```

Before the loop be sure to create the space for `cdfs` by setting `cdfs=vector(,trials)`. For each trial, store the estimated value of the CDF at the point 0. Calculate the variance of the estimate $F_n(0)$ and the $bias(F_n(0))$ across the 1000 simulations.

d. Now repeat part c. but with a sample size of 500. Compare the mean squared error across the two different sample sizes.