

Lab E/F - SOCI709

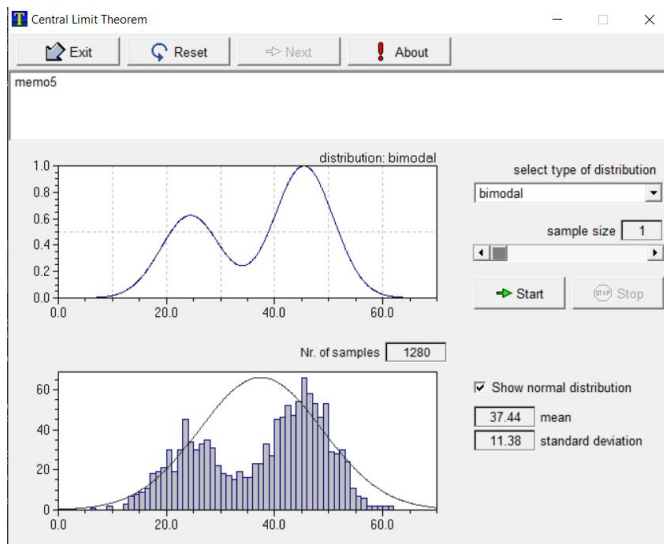
Kate Brandt

February 10, 2019

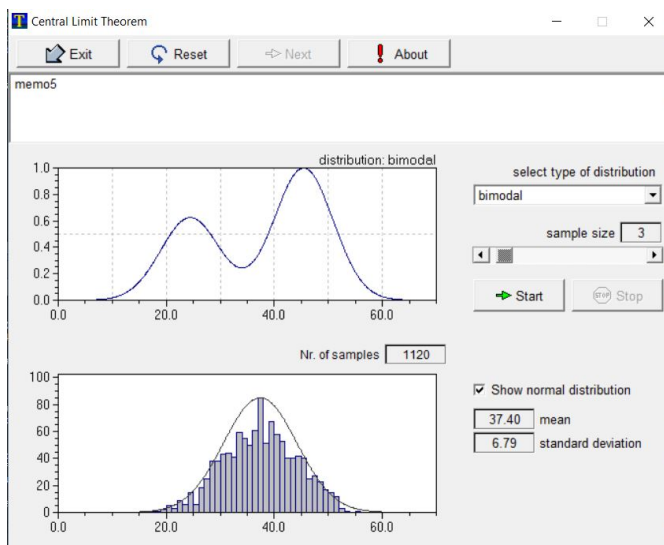
Q1. What size N seems to make the bimodal distribution look normal?

- The greater the N, the more normal the distribution appears. The three images below show how the distribution becomes normal as N increases from 1 to 3 to 6. An N of 1 most prominently shows the bimodal distribution because each sample's average equals the one sample. As the N increases, the distribution of each sample is the average of the N samples, so the distribution is distributed around the mean of the two modes. By about $N = 5$ or $N = 6$, the bimodal distribution appears normal.

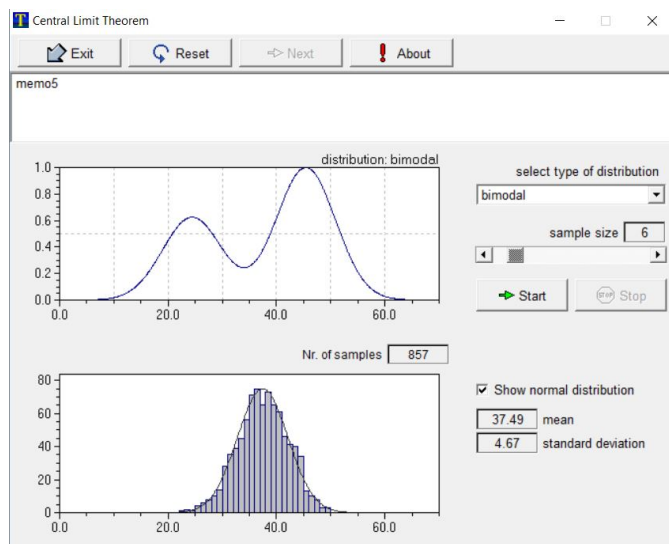
N = 1



N = 3



N = 6



Q2. Monte Carlo simulation of the variance b

For this part of the assignment, I used the MonteCarlo package in R. The link which explains the functionality of this package can be found [here](#)

```
#Setting up
#Loading package
library(MonteCarlo)
library(dplyr)
library(tinytex)

#The following recreates the estols2.ado file
estols2 <- function(n, xmean=0, x_sd=3, emean=0, e_sd=15, xcoef=3){

# The following steps are based on Lab E STATA code
# Create data for x, y, and e

# Create sample x [from STATA: "gen x=3*invnorm(uniform())"]
#(standard deviation = 3, mean = 0)
x <- rnorm(n, mean = xmean, sd = x_sd)

#Create sample e [from STATA: "gen e=15*invnorm(uniform())"]
e <- rnorm(n, mean = emean, sd = e_sd)

#Create sample y
y <- xcoef*x+e

#create data frame to regress
df <- data.frame(y,x,e)

#output regression
model <- (lm(y~x, data = df))
coeflist <- model$coefficients[2] %>% unname

return(list("coeflist" = coeflist))
}
```

```

# Use the estols2 function to run Monte Carlo simulations with different parameters
# MonteCarlo takes 3 arguments: function, reps, and grid of parameters for function
# Define parameter grid
n_grid <- 300
param <- list("n"=n_grid)
estols2(n_grid)
MC_q2_a <- MonteCarlo(func = estols2, nrep = 200, param_list = param)

```

```

#Summarize output with table and graph
library(ggplot2)
MC_2a <- MakeFrame(MC_q2_a)
tbl_2a <- tbl_df(MC_2a)
coef2a <- tbl_2a$coeflist
summary(coef2a)

```

```

##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      2.139   2.800   2.969   2.987   3.188   3.738

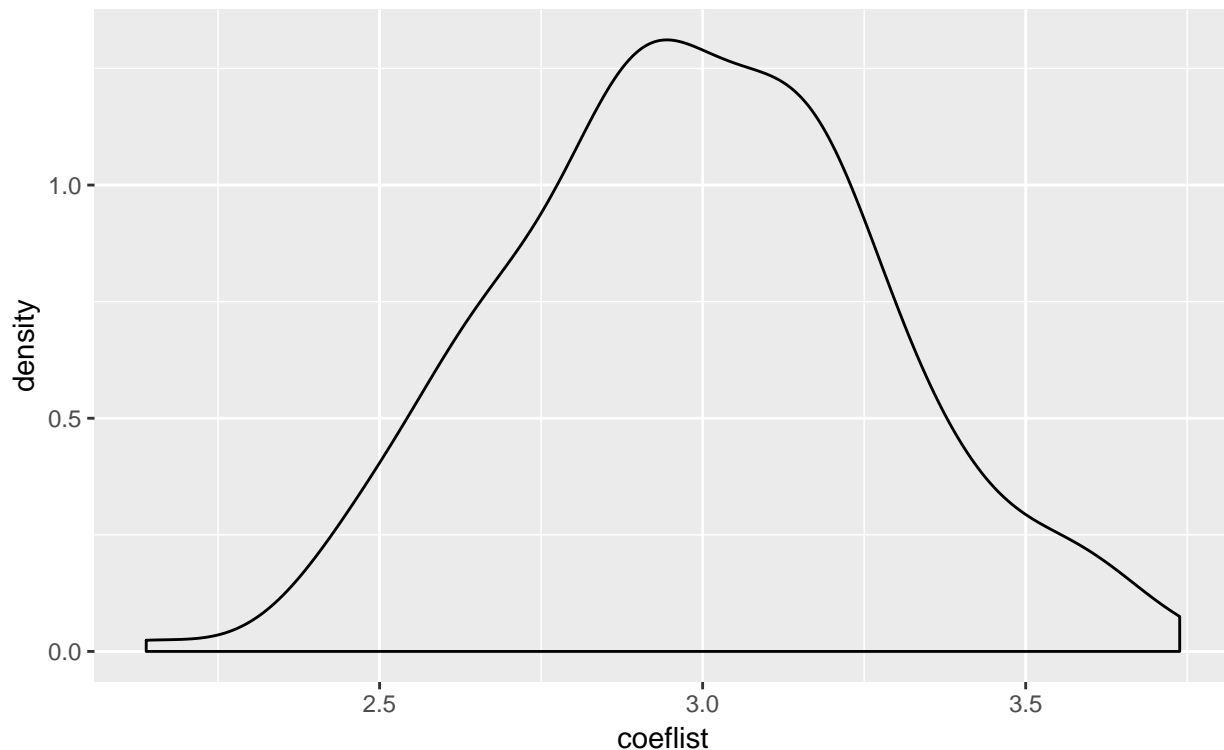
```

```
sd(coef2a)
```

```
## [1] 0.2871605
```

(Q2a) Observed Distribution of B

N=300, reps=200, normal error distribution (based on estols2 parameters)



Q2, Part B: N=30, reps=300

```

n_grid <- 30
param <- list("n"=n_grid)
estols2(n_grid)
MC_q2_b <- MonteCarlo(func = estols2, nrep = 300, param_list = param)

```

```
#Summarize output with table and graph
```

```
library(ggplot2)
MC_2b <- MakeFrame(MC_q2_b)
tbl_2b <- tbl_df(MC_2b)
coef2b <- tbl_2b$coeflist
summary(coef2b)
```

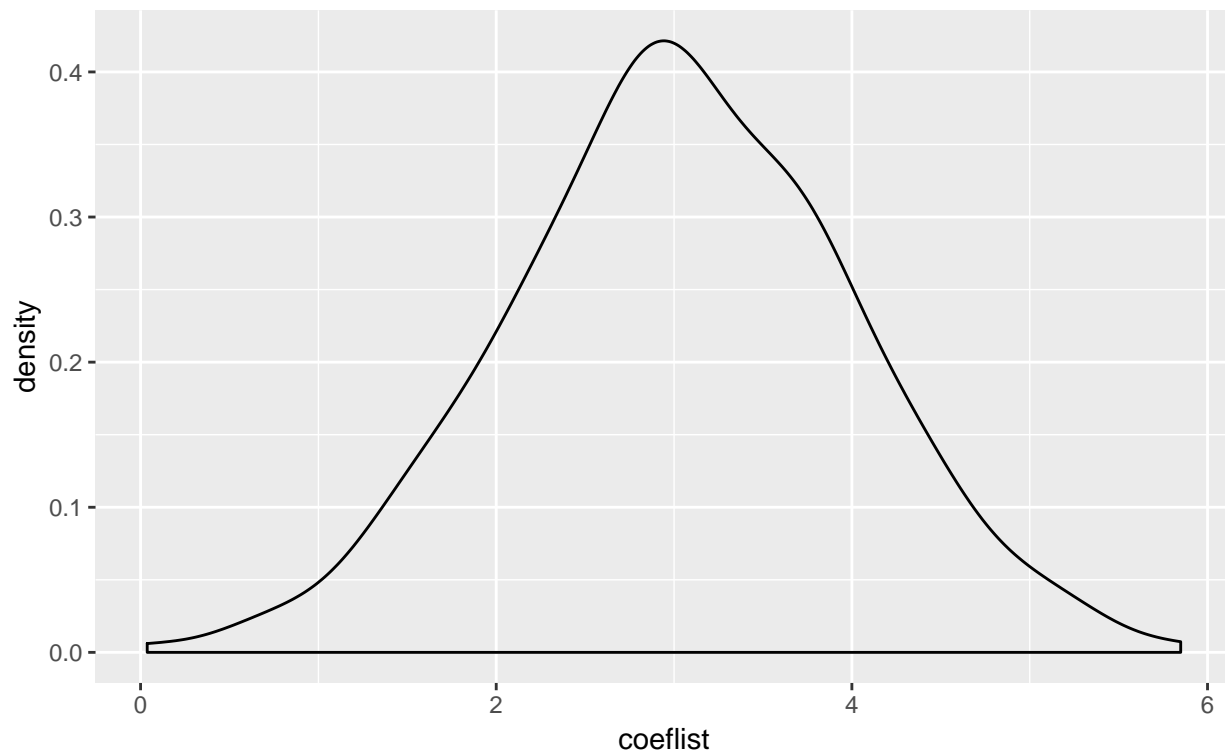
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.03727 2.40832 3.00553 3.03650 3.67877 5.84917
```

```
sd(coef2b)
```

```
## [1] 0.9636563
```

(Q2b) Observed Distribution of B

N=30, reps=300, normal error distribution (based on estols2 parameters)



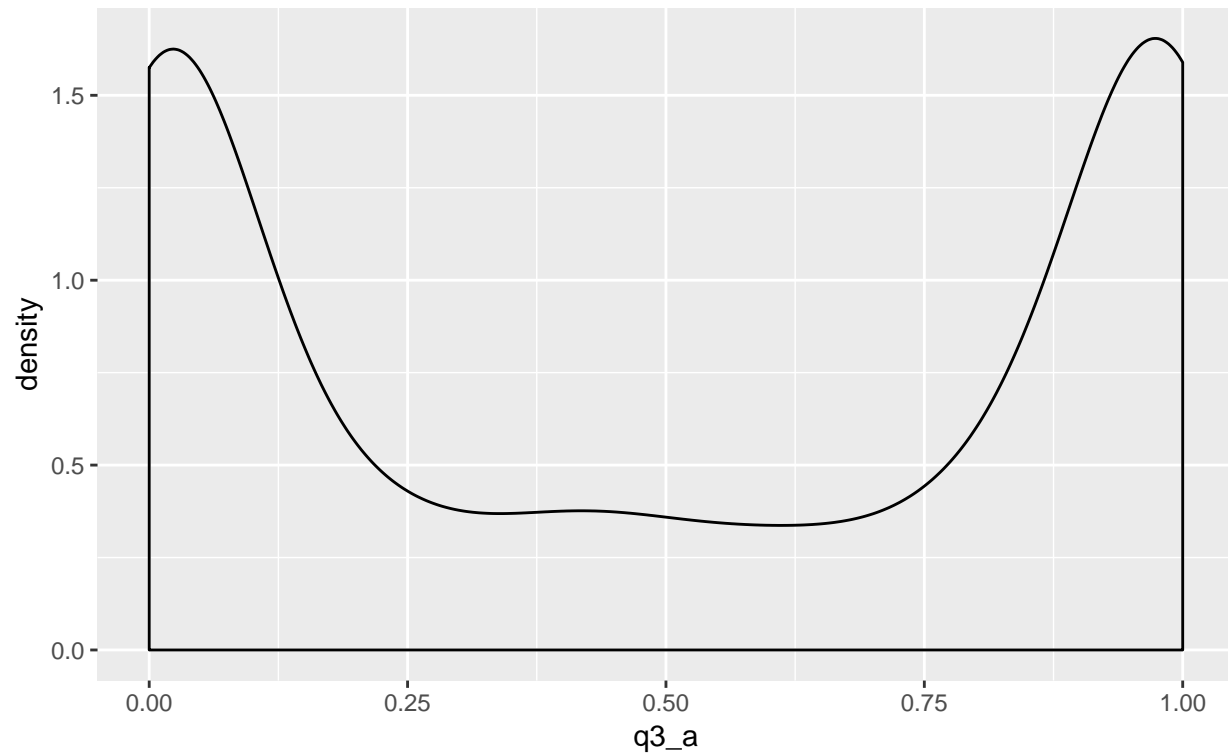
Q3.Simulating different error terms

The following is based on recreating the test_dist.do

```
# Using rbeta() command to simulate beta distribution. Previous problems were using rnorm() for normal
# obs = 2000
q3_a <- rbeta(2000, 0.2, 0.2)
```

(Q3a) Observed Distribution of B

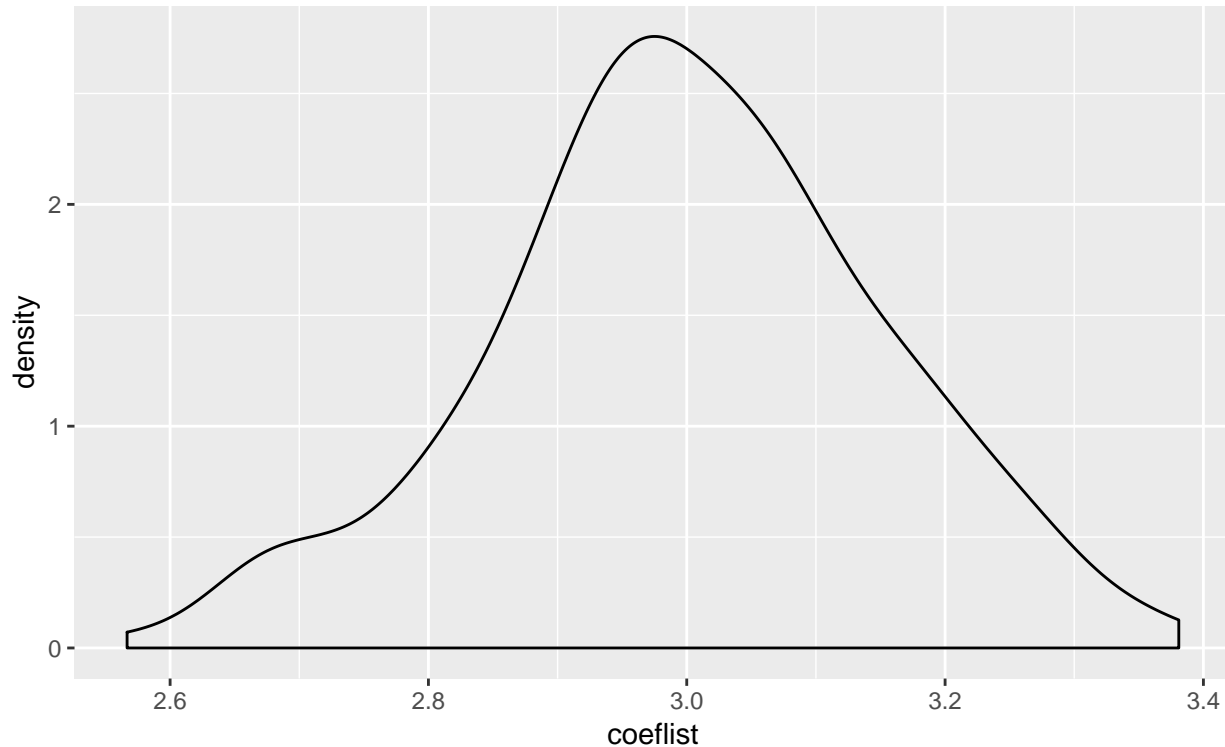
reps=2000, bimodal distribution ($\alpha = 0.2$, $\beta = 0.2$)



Substitute this command for estols2.ado (from question 2) This will end up being written in the syntax as:
`e_q3 <- 15*rbeta(n, shape1 = 0.2, shape2 = 0.2)`

(Q3b) Observed Distribution of B

N=200, reps=200, beta distribution of error term ($\alpha = 0.2$, $\beta = 0.2$)



This beta error term distribution, using the simulated data samples, seems to result in a slightly abnormal distribution of b , but not one that is much more noticeable than the abnormalities from the normal distributions in Q2. This is perhaps because of the high number of N and high number of reps. Even abnormal distributions, repeated enough times and added together, can appear normal. In experimenting with the distribution module from question 1, it seems as though an N as small as 5 can begin result in a seemingly even distribution as the observations from each experiment average out. In part A of this question, the beta distribution experiment was conducted 2000 times, but with one observation per rep. This is what allowed the distribution to appear bimodal. This is also why, with 200 reps and $N=200$ in part B, the distribution appears mostly normal.

Q4. How do these simulations constitute an “experiment” of the validity of asymptotic normality of the distribution of b ?

All of these simulations create sets of randomly generated values from a specified distribution. From these sets, each “experiment” is conducted by a random selection from the data set to be analyzed (in this case, regressed). Each regression produces a coefficient, b , which describes the relationship between y and x . Calculations of these coefficients, repeatedly, generate a set of estimated “ B ” values which can be plotted to determine density of the estimations. This is an experiment because of the randomness of the data and selected samples, and the repetitions which accumulate a mass of data points. Though these are simulated data, the randomness of the set error term (via normal or beta density) allow for variation. This is why we plot the estimated values: to assess the effect of the error term.