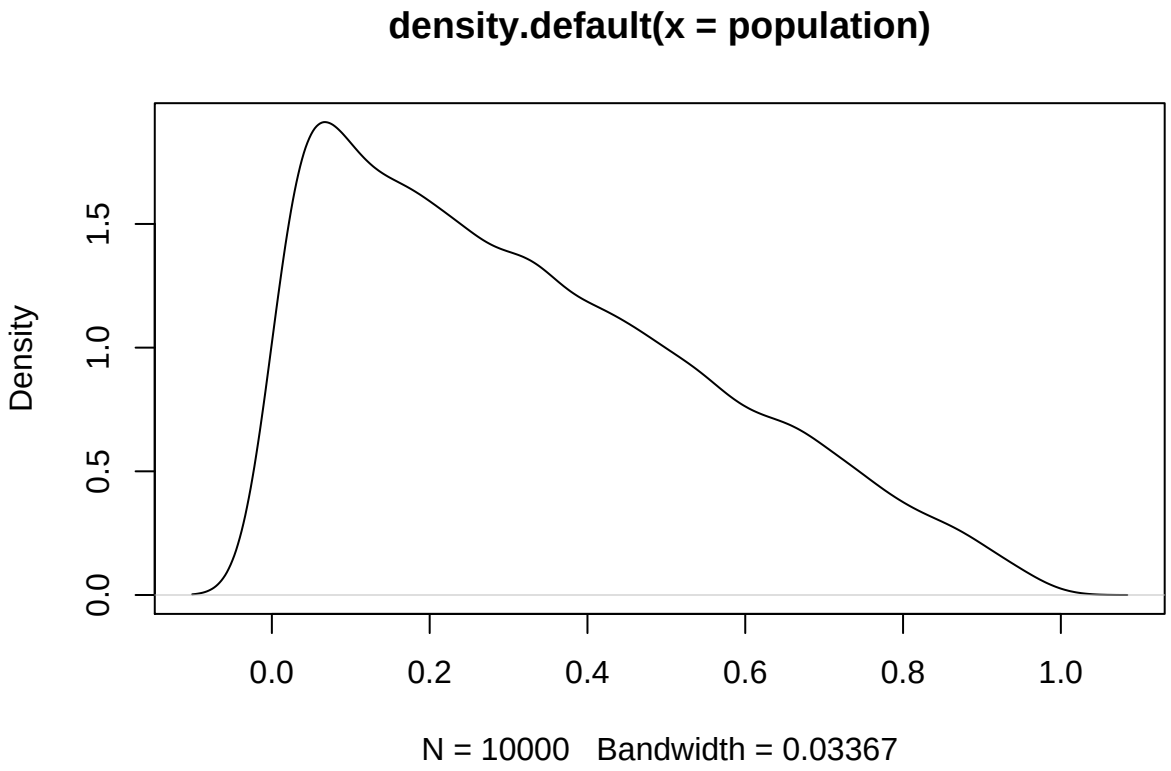
Lecture 1

Han Zhang

2/1/2021

Population is 10000 units following Beta distribution.

```
population <- rbeta (10000, 1, 2)
plot (density (population))
```




```
# population mean
mean(population)
```

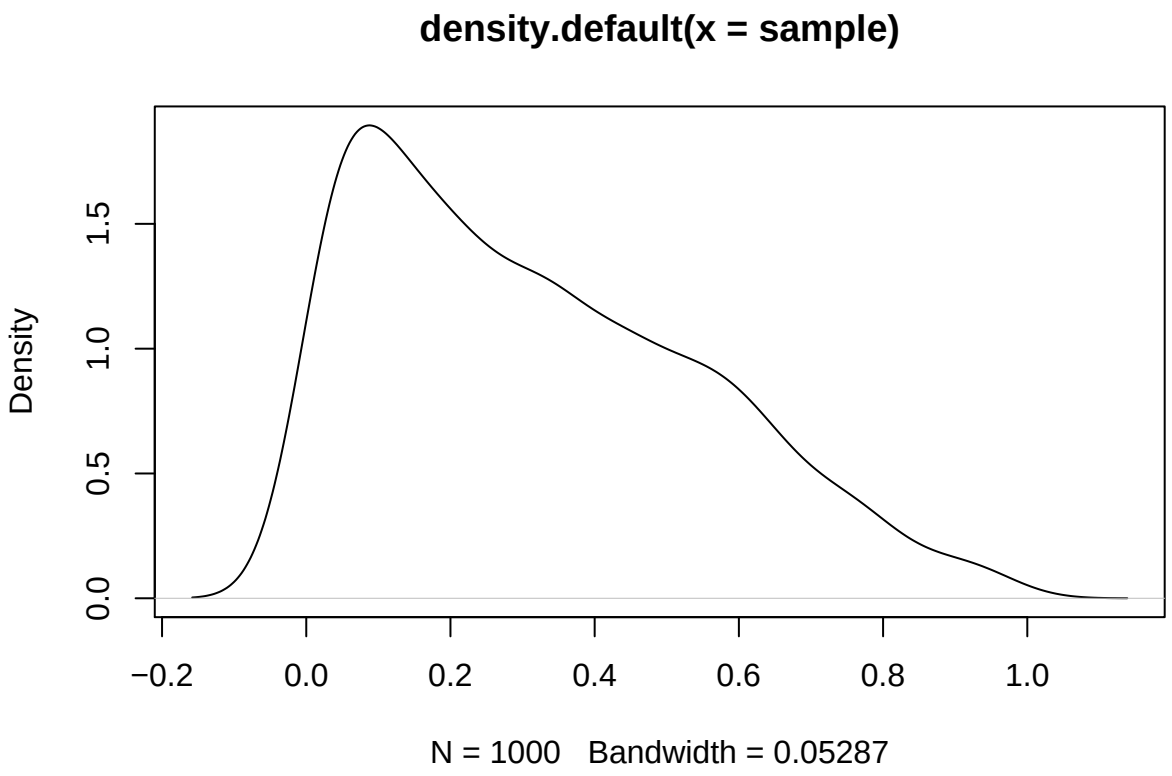
```
## [1] 0.3294625
```

```
# population variance
var(population)
```

```
## [1] 0.05573201
```

Create samples from the population with size 1000

```
sample = sample(population, size = 1000)
plot (density (sample))
```




```
# sample mean
mean(sample)
```

```
## [1] 0.3182681
```

```
# sample variance of the sample mean
var(population) / length(sample)
```

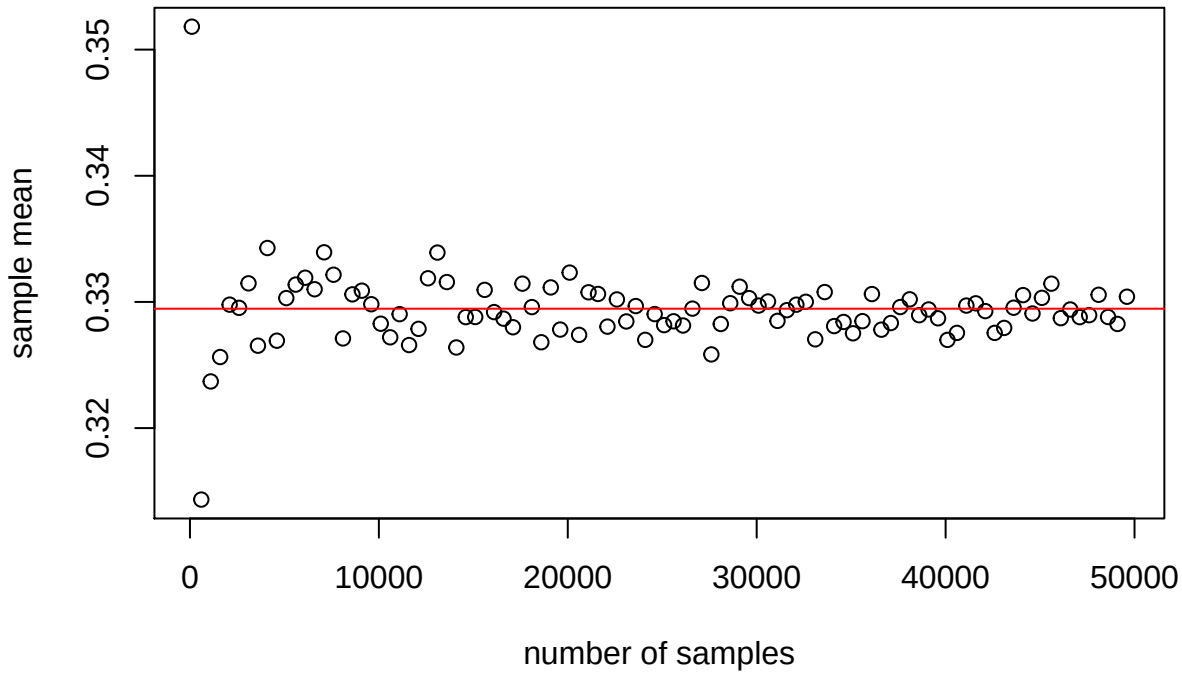
```
## [1] 5.573201e-05
```

#$ Law of Large Numbers

As n increases, sample mean approaches population mean.

```
sample_mean_list <- c()
sample_times <- seq(100, 50000, 500)
for (n in sample_times ){
  sample <- sample(population, size = n, replace = TRUE)
  sample_mean_list <- c(sample_mean_list, mean(sample))
}

plot(sample_times, sample_mean_list, xlab = "number of samples", ylab = "sample mean")
abline(h = mean(population), col = "red")
```



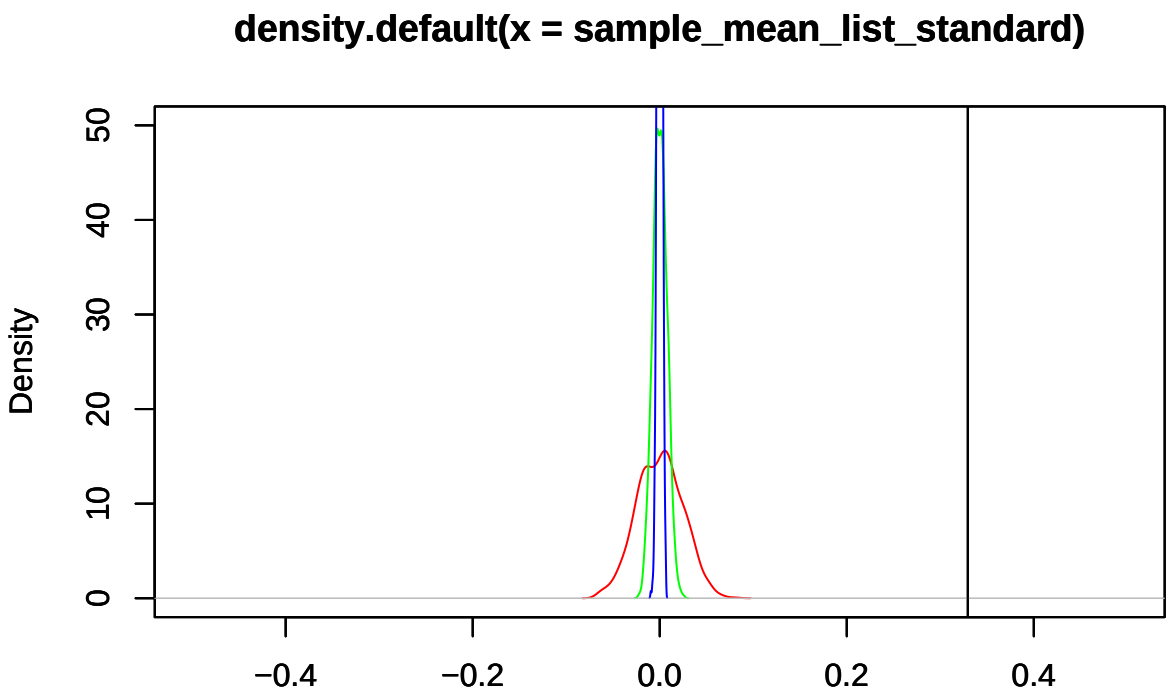
#$ Central limit theorem - As n increases, distribution of sample means approaches normal

```
# Now we show the central limit theorem
sample_times <- c(100, 1000, 10000)
col <- c("red", "green", "blue")
i = 1
for (n in sample_times ){
  # central limit theorem talks about distribution of the sample mean
  # we cannot calculate the distribution for a single sample, so we draw sample multiple times
  sample_mean_list <- c()
  for (m in 1:1000)
  {
    sample <- sample(population, size = n, replace = TRUE)
    sample_mean_list <- c(sample_mean_list, mean(sample))
  }
  sample_mean_list_standard <- sample_mean_list - mean(population)
  plot(density (sample_mean_list_standard),  col = col[i], xlim = c(-0.5, 0.5), ylim  = c(0,50), xlab =
  abline(v = mean(population), col = "black")
  par(new = T)
  i =i  + 1
}

# bootstrap -----------------



# print (paste ("sample size is ", n, "sample mean is ", mean(sample)))
```


density.default(x = sample_mean_list_standard)
Density
0
10
20
30
40
50
−0.4
−0.2
0.0
0.2
0.4