ASRM569 HW1

Minhao Zhang

September 2020

Question One

Proof:
1. The more samples we take, the more likely that the sampling distribution of the means will be normally distributed.
2. The larger the sample, the tighter the sample means will tend to be around the population mean.

Proof by R

In order to prove the above-mentioned properties, I simulated four distinct groups of random values conforming with the exponential function (rate=0.5). I respectively sampled 1/2/5/100 observations each trial and conducted 2500 trails for each group.After plotting the corresponding four histograms, I fitted the normal distributions on the histograms with a blue line to see whether there were good fits. (github link: https://github.com/zhangminhao00/569HW.git)

```r
set.seed(569)
sample_size = c(1,2,5,100)
n_samples = 2500
means=c()
std=c()
variance=c()

par(mfrow=c(1,4))
for (i in c(1:4)){
simulated_number_accident = rexp(sample_size[i]*n_samples,rate=0.5)
simulated_matrix = matrix(simulated_number_accident, n_samples)
sample_means = rowMeans(simulated_matrix)
means=c(means,mean(sample_means))
std=c(std,sd(sample_means))
variance=c(variance,sd(sample_means)^2)


main=c("Means of Exp(rate=0.5) \n sample size:\n",sample_size[i])
hist_exp = hist(sample_means,xlab="means",
                main=main,
                ylim=c(0,850))
xfit_exp = seq(min(sample_means),max(sample_means),length=100)
yfit_exp = dnorm(xfit_exp,mean=mean(sample_means),sd=sd(sample_means))
yfit_exp_scaled = yfit_exp*diff( hist_exp$mids[1:2])*length(sample_means)
lines(xfit_exp, yfit_exp_scaled, col="blue", lwd=2)}

for (i in c(1:4)){
cat("\n\n When whe sample size is",sample_size[i],"\n the mean is", means[i],
    "\n the standard deviation is",std[i],"\n the variance is",variance[i])}
```

Figure 1: Codes

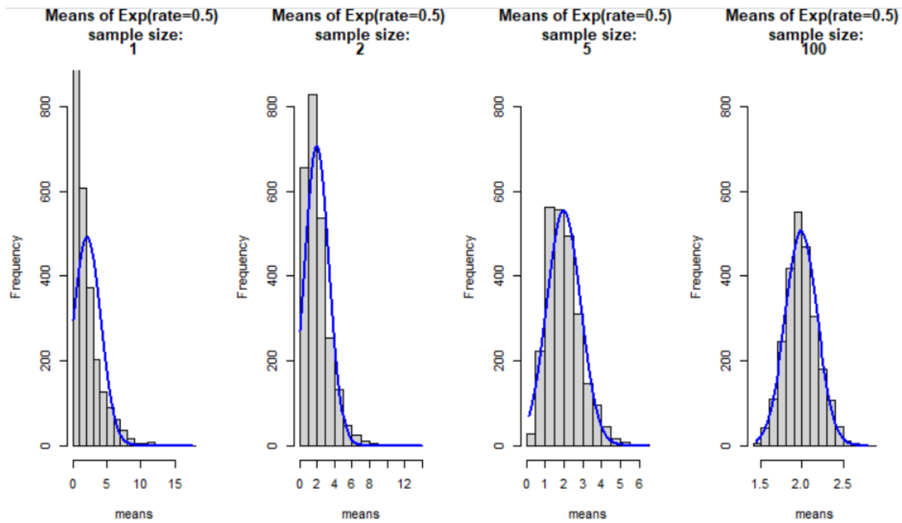

Figure 2: CLT

As we can see in the picture, when the sample size increases, the sampling distribution of the means has a greater fit with the blue line, which means the sampling distribution of the means becomes more normally distributed. In addition, the range of x-axis coordinates is much narrower with large sample size, so we can conclude that the sample means are tighter around the population mean as the sample size increases.

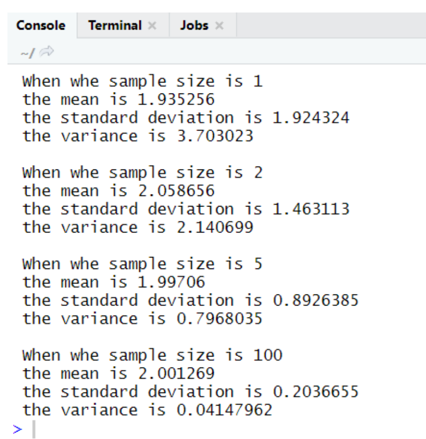

Figure 3: Outcomes

Furthermore, the calculation of the means and variances helps us have a better understanding of the second property. As the sample size increases, the standard deviations and the variances significantly decrease, which means the sample means are tighter around the population mean.

Question Two

What is the difference between CLT and law of large numbers?

Answer:
The Central limit Theorem(CLT) states that the sampling distribution of the samples' means approaches a normal distribution as the samples' sizes get larger. The law of large numbers(LLN) states that the average of the results obtained from a large number of trials should be close to the expected value and will tend to become closer to the expected value as more trials are performed.
In another word, CLT states that when sample size tends to infinity, the sample mean will be normally distributed, while LLN states that when sample size tends to infinity, the sample mean equals to population mean.
More briefly, CLT is a statement about the SHAPE of the distribution. LLN tells us where the center of the bell is located.
To be specific, in the case study conducted in Question One, as the sample size increases, CLT tells us that the sampling distribution of the means will be more likely to be normally distributed and LLN points out that the center of the normal distribution will be more likely to be 2.