Statistical Inference Course Project: Part I

Yifan XIA

4 Feb 2017

Simulation exercise

Overview

A sampling process from an exponential distribution is investigated in this report. We perform 1000 simulations, with 40 samples per simulation. The sample mean distribution of all simualtions is then compared with the exponenial distribution in R using the Central Limit Theorem.

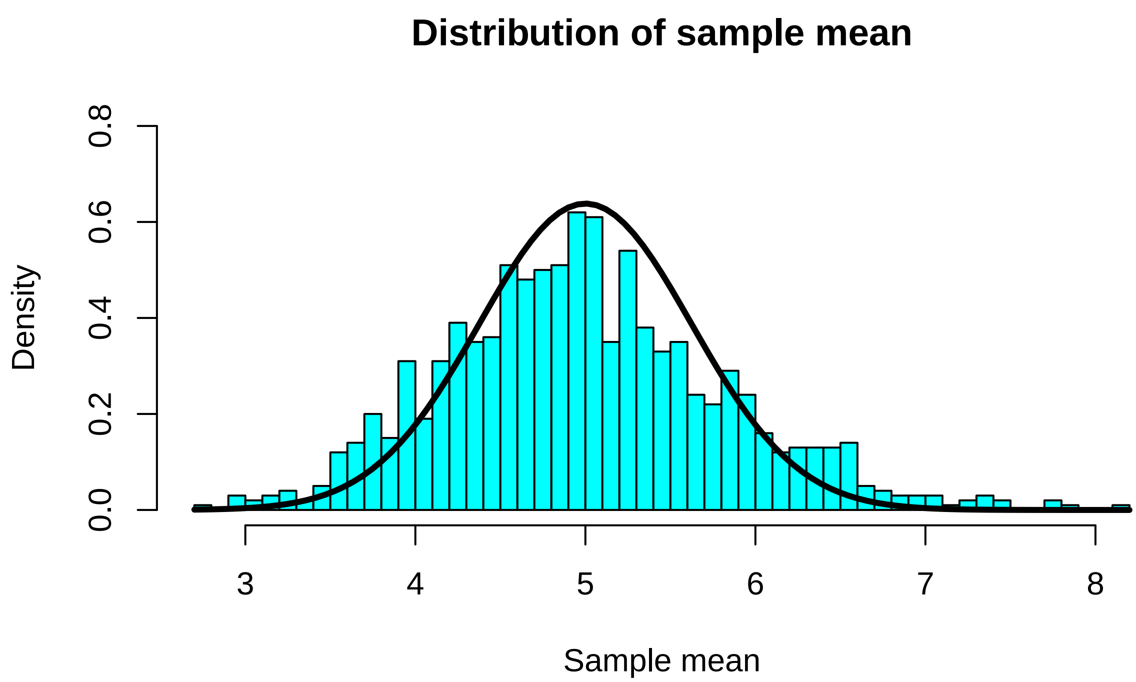
Run simulation

```
lambda <- 0.2; nExps <- 40; nSimu <- 1000
mns <- NULL
for (i in 1:nSimu) mns <- c(mns, mean(rexp(nExps, lambda)))
sampleMean <- mean(mns); sdErr <- sd(mns)/sqrt(nExps)
```

Compare simulated distribution and a normal distribution

We first visually compare the simulated distribution with a normal distribution resulted from an exponential distribution of $\lambda = 0.2$:

```
hist(mns, breaks = 50, prob = TRUE, col = 'cyan', xlab = 'Sample mean',
     ylab = 'Density', ylim = range(0,0.8), main = 'Distribution of sample mean')
curve(dnorm(x, mean=1/lambda, sd=1/lambda^2/nExps), add=TRUE, col='black', lwd = 3)
```



The above figure shows a histogram of sample mean (cyan), which is compared with the probability density function of the normal distribution (black curve):

$$\mathcal{N}(\mu = \frac{1}{\lambda}, \quad \frac{\sigma^2}{n} = \frac{1}{n\lambda^2}),$$

where λ=0.2 and n=40. The comparison suggests that the sample mean follows a normal distribution given in the above equation. Now let us compare the sample mean and the variance with their theoretical values.

```
samples <- c(1/lambda, sampleMean, (1/lambda^2)/nExps, sd(mns)^2)
names(samples) <- c('True mean','Sample mean','True variance', 'Sample variance')
print(samples)
```

```
##       True mean     Sample mean   True variance Sample variance
##       5.0000000       4.9898187       0.6250000       0.6762999
```

The above output shows that the sample mean and the true mean are quite close, as well as the true variance and the sample variance.

Another important aspect of a distribution simulation is its confidence intervals. We need to verify the inclusion of the true mean in the simulated confidence interval. The simulated 95% confidence interval is computed by the following code:

```
confInt <- sampleMean + c(-1,1)*qnorm(0.975)*sdErr
confInt
```

```
## [1] 4.734967 5.244670
```

The above confidence interval includes the true mean $1/\lambda$=5.

The distribution of the sample means can be approximated to a normal distribution $\mathcal{N}(\overline{X}, \sigma_X)$, where $\overline{X}$ denotes the average of the sample means, and σ_X denotes the samples' standard deviation. The probability function of this normal distribution is plotted below, together with the histogram of the sample means.

```
hist(mns, breaks = 50, prob = TRUE, col = 'cyan', xlab = 'Sample mean',
     ylab = 'Probability', main = 'Distribution of sample mean')
curve(dnorm(x, mean=sampleMean, sd=sd(mns)), add=TRUE, col='red', lwd = 3)
```