

Statistical Inference course project - Simulation exercises

Wei-Meng Lee

First, simulate 1,000 averages of 40 (pseudo-)random samples from the exponential distribution, with lambda (the rate parameter) set to 0.2. Also, for the purpose of reproducibility of the graphs seen in this report, we set a seed.

```
set.seed(1)
n <- 40
trials <- 1000
lambda <- 0.2
simulation <- sapply(rep(1, trials), function(x) x * mean(rexp(n, lambda)))
```

Question 1

Show where the distribution is centered at and compare it to the theoretical center of the distribution.

The center of distribution for our simulation is simply the median:

```
median(simulation)
```

```
## [1] 4.95
```

This is approximate to the theoretical center of the distribution (according to the central limit theorem, the mean of the sample means is the same as the population mean):

```
theoretical.mean <- 1/lambda
theoretical.mean
```

```
## [1] 5
```

Question 2

Show how variable it is and compare it to the theoretical variance of the distribution.

The variance of our simulation is:

```
var(simulation)
```

```
## [1] 0.6111
```

This is approximate to the theoretical variance of the distribution according to the central limit theorem:

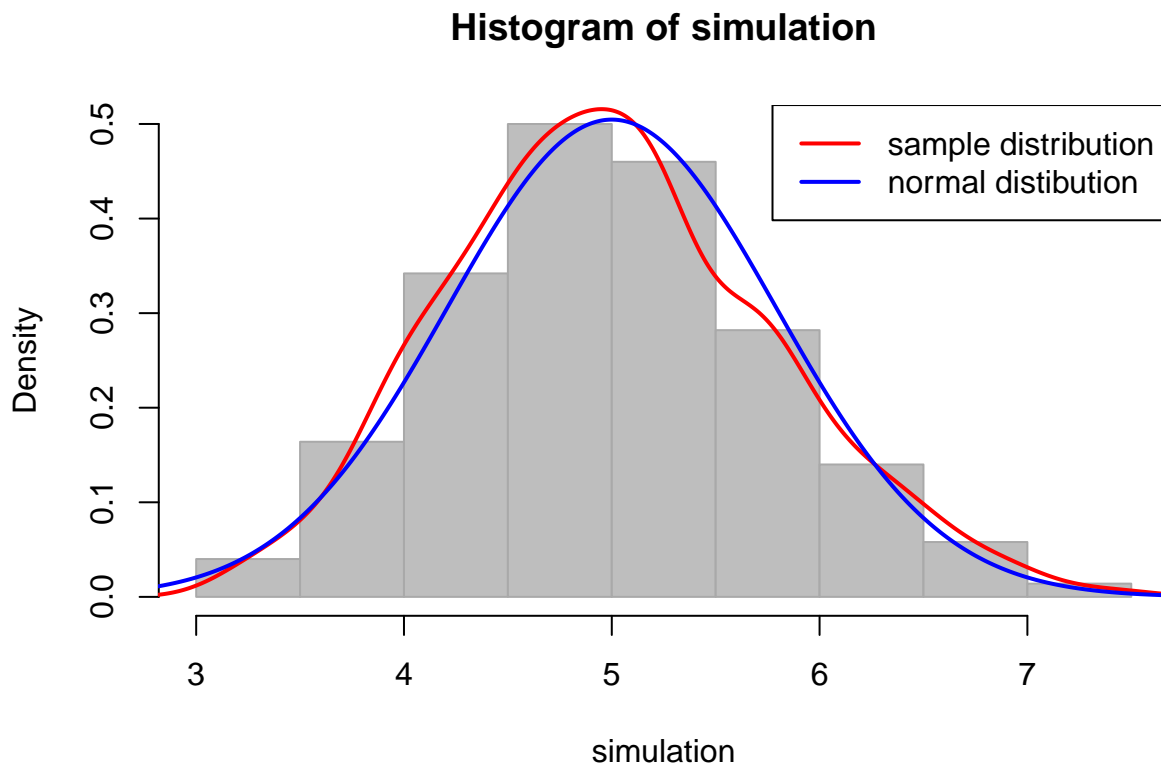
```
theoretical.variance <- 1/n * (1/lambda)^2
theoretical.variance
```

```
## [1] 0.625
```

Question 3

Show that the distribution is approximately normal.

```
hist(simulation, prob=TRUE, col="grey", border="dark grey")
lines(density(simulation), lwd=2, col="red")
# Plot a normal distribution using theoretical parameters
x.norm <- seq(min(density(simulation)$x), max(density(simulation)$x), length=1000)
y.norm <- dnorm(x.norm, mean = theoretical.mean, sd = sqrt(theoretical.variance))
lines(x.norm, y.norm, lwd=2, col="blue")
legend("topright", c("sample distribution", "normal distribution"), col = c("red", "blue"), lwd=2)
```



It can be observed that the sample distribution approximates a normal distribution.