Statistical Inference course project - Simulation exercises

Wei-Meng Lee

- Simulate 1,000 averages of 40 (pseudo-)random samples from the exponential distribution, with lambda (the rate parameter) set to 0.2.
- Set a seed so graphs in this report are reproducible.

```
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 thereotical variance of the distribution according to the central limit theorem:

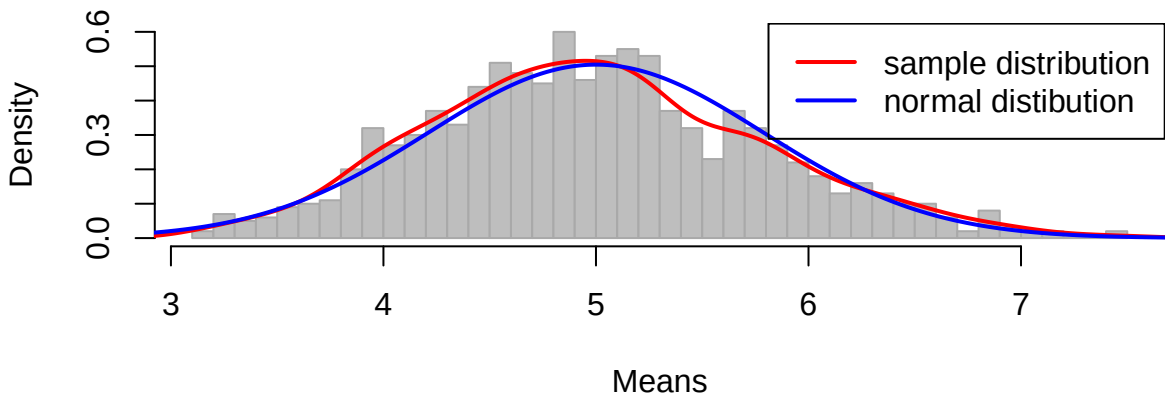
```
theoretical.variance <- 1/n * (1/lambda)^2; theoretical.variance
```

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

Question 3

> Show that the distribution is approximately normal.

```r
hist(simulation, prob=TRUE, breaks=50, col="grey", border="dark grey", main = NULL, xlab = "Means")
lines(density(simulation), lwd=2, col="red")
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 distibution"), col = c("red", "blue"), lwd=2)
```

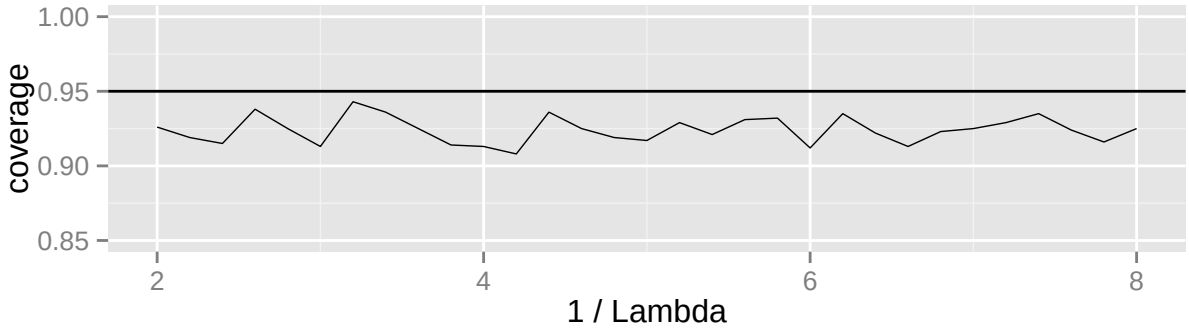


It can be observed that the sample distribution approximates a normal distribution.

Question 4

Evaluate the coverage of the confidence interval for 1/lambda.

```r
meanvals <- seq(2, 8, by = 0.2)
coverage <- sapply(meanvals, function(meanval) {
        simulation <- sapply(rep(1, trials), function(x) x * (rexp(n, 1/meanval)))
        mean.mean <- apply(simulation, 2, mean); mean.sd <- apply(simulation, 2, sd)
        ll <- mean.mean - qnorm(.975) * mean.sd / sqrt(n)
        ul <- mean.mean + qnorm(.975) * mean.sd / sqrt(n)
        mean(ll < meanval & ul > meanval)
})
library(ggplot2)
ggplot(data.frame(meanvals, coverage), aes(x = meanvals, y = coverage)) + geom_line(size=0.25) +
        geom_hline(yintercept=0.95, size = 0.5) + ylim(.85, 1.0) + xlab("1 / Lambda")
```



It can be observed that coverage is about 92.5% of the 95% confidence interval, with the value of `1 / lambda` not affecting the coverage.