Statistical Inference Course Project - Part 1

John Letteboer

10/21/2018

Synopsis

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. A simulation exercise.
2. Basic inferential data analysis.

Simulation

The exponential distribution can be simulated in R with `rexp(n, lambda)` where lambda is the rate parameter. The mean of exponential distribution is **1/lambda** and the standard deviation is also **1/lambda**. **Set lambda = 0.2 for all of the simulations**. You will investigate the distribution of averages of **40 exponentials**. Note that you will need to do a **thousand simulations**.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1. Show the sample mean and compare it to the theoretical mean of the distribution.

2. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

3. Show that the distribution is approximately normal.

What parameters do we have:
- lambda of 0.2
- 40 exponentials
- thousand (1000) simulations

```
# Set seed for reproducibility
set.seed(1234)

# lambda of 0.2
lambda <- 0.2
# 40 exponentials
n <- 40
# thousand (1000) simulations
simulations <- 1000

# simulation of exponentials
sim_exps <- replicate(simulations, rexp(n, lambda))

# mean of simulation
mean_exps <- apply(sim_exps, 2, mean)
```

Question 1

Show the sample mean and compare it to the theoretical mean of the distribution.

The center of mass of the data is the empirical mean $\bar{X} = \sum_{i=1}^{n} x_i p(x_i)$ where $p(x_i) = 1/n$

```
# sample mean
sample_mean <- mean(mean_exps)
sample_mean
```

```
## [1] 4.974239
```

```
# theoretical mean
theo_mean <- 1/lambda
theo_mean
```

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

Results
The sample mean is **4.9742388**, the theoretical mean **5**. The center of distribution of averages of 40 exponentials is very close to the theoretical center of the distribution.

Question 2

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

```
# standard deviation of distribution
sd_dist <- sd(mean_exps)
sd_dist
```

```
## [1] 0.7554171
```

```
# standard deviation of theoretical expression
sd_theo <- 1/lambda/sqrt(n)
sd_theo
```

```
## [1] 0.7905694
```

```
# variance of distrubution
var_dist <- sd_dist^2
var_dist
```

```
## [1] 0.5706551
```

```
# variance of theoretical expression
var_theo <- sd_theo^2
var_theo
```

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

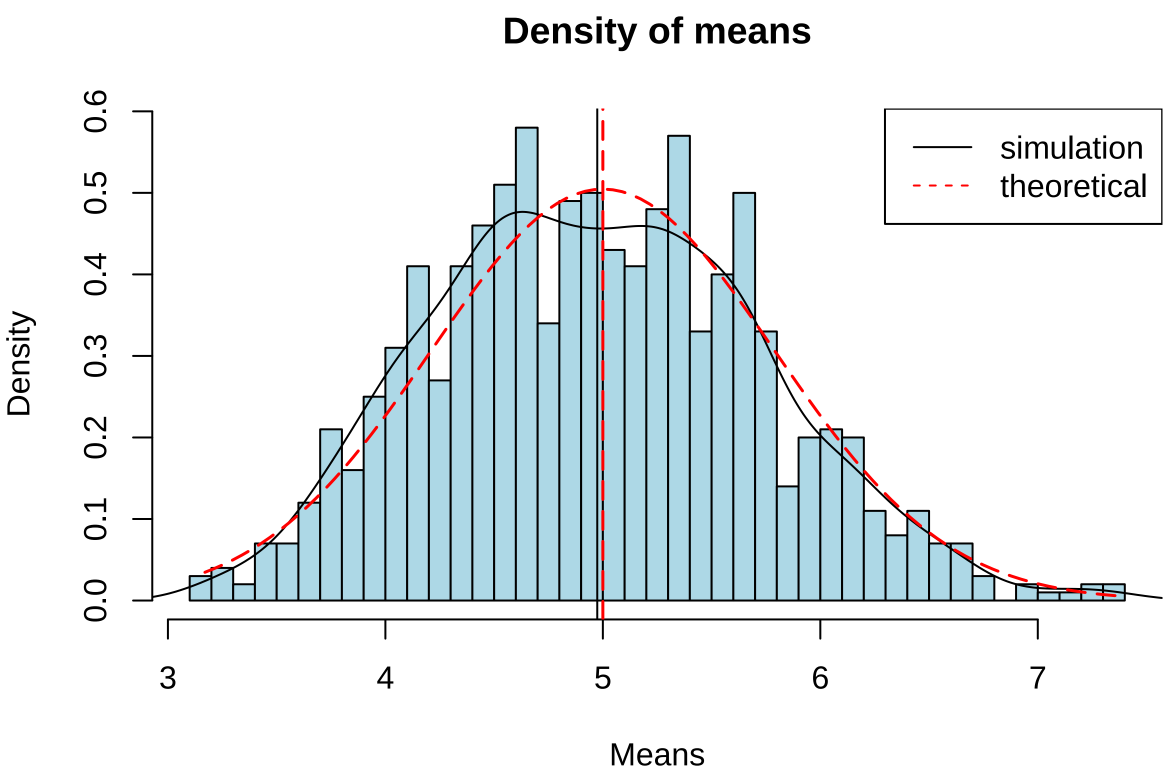
Results
Standard Deviation of the distribution is 0.7554171 with the theoretical SD calculated as 0.7905694. The Theoretical variance is calculated as $(\frac{1}{\lambda}\frac{1}{\sqrt{n}})^2 = 0.625$. The actual variance of the distribution is 0.5706551

Question 3

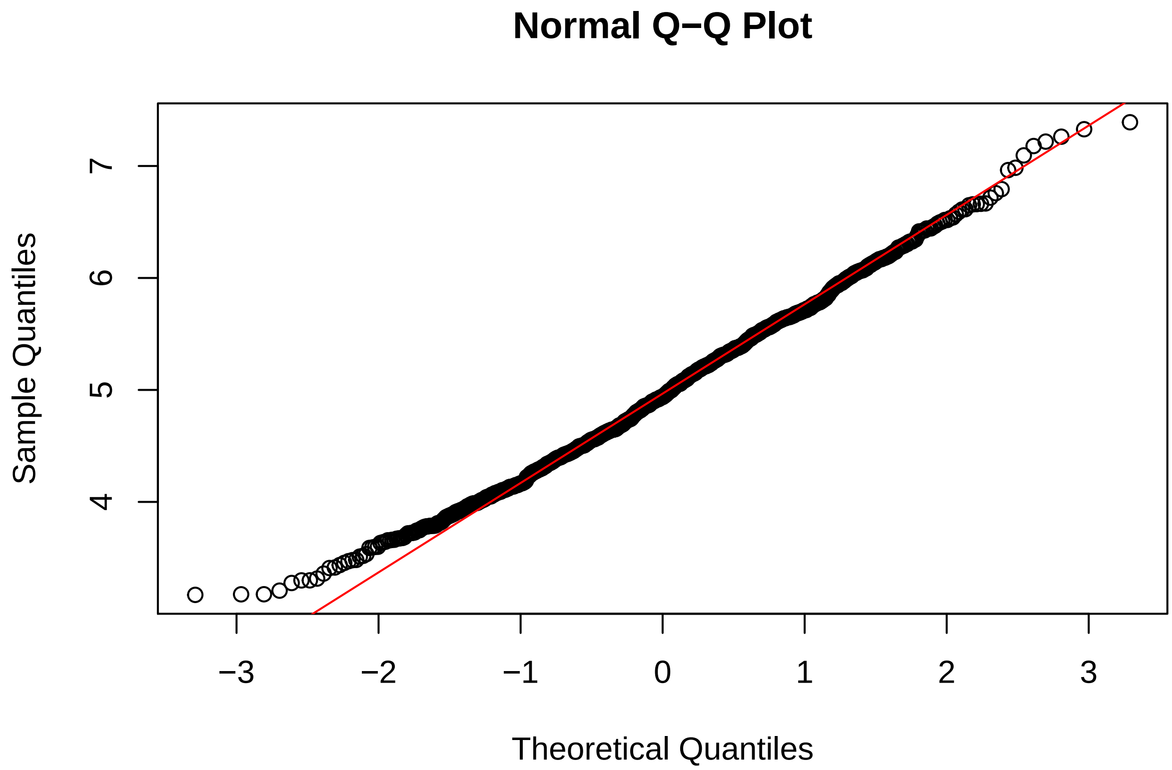
Show that the distribution is approximately normal.

```
xfit <- seq(min(mean_exps), max(mean_exps), length=100)
yfit <- dnorm(xfit, mean=theo_mean, sd=sd_theo)
hist(mean_exps,
     breaks=n,
     prob = T,
     col ="lightblue",
     xlab = "Means",
     ylab = "Density",
     main = "Density of means")

# Add Sample mean
abline(v=sample_mean)
# Add density of the averages of samples
lines(density(mean_exps))
# Add normal curve
lines(xfit, yfit, lwd=1.5, pch=22, col="red", lty=5)
# Add theoretical center of distribution
abline(v=theo_mean, lwd=1.5, pch=22, col="red", lty=5)
# Add legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "red"))
```




```
# compare the distribution to a normal distribution
qqnorm(mean_exps)
qqline(mean_exps, col = 2)
```



Results

As you can see from the graphs, the calculated distribution of means of random sampled exponantial distributions, overlaps quite nice with the normal distribution with the expected values based on the given lamba