Statistical Inference - Course Project Part 1

Author : Davy Meesemaecker

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.

Simulation

Let's start by creating our parameters :

```
set.seed(13053312)
lambda <- 0.2
n <- 40
sim <- 1000
```

Now we can start our simulation :

```
sim_test <- matrix(data=rexp(n * sim, lambda), sim)
sim_test_Means <- apply(sim_test, 1, mean)
```

```
str(sim_test)
```

```
##  num [1:1000, 1:40] 10.54 1.91 16.46 2.68 2.84 ...
```

Sample Mean vs Theoretical Mean

What is our sample mean ?

```
rexp_sim_mean <- mean(sim_test_Means)
rexp_sim_mean
```

```
## [1] 4.985525
```

How about our theoretical mean ?

```
theoretical_mean <- 1/0.2
theoretical_mean
```

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

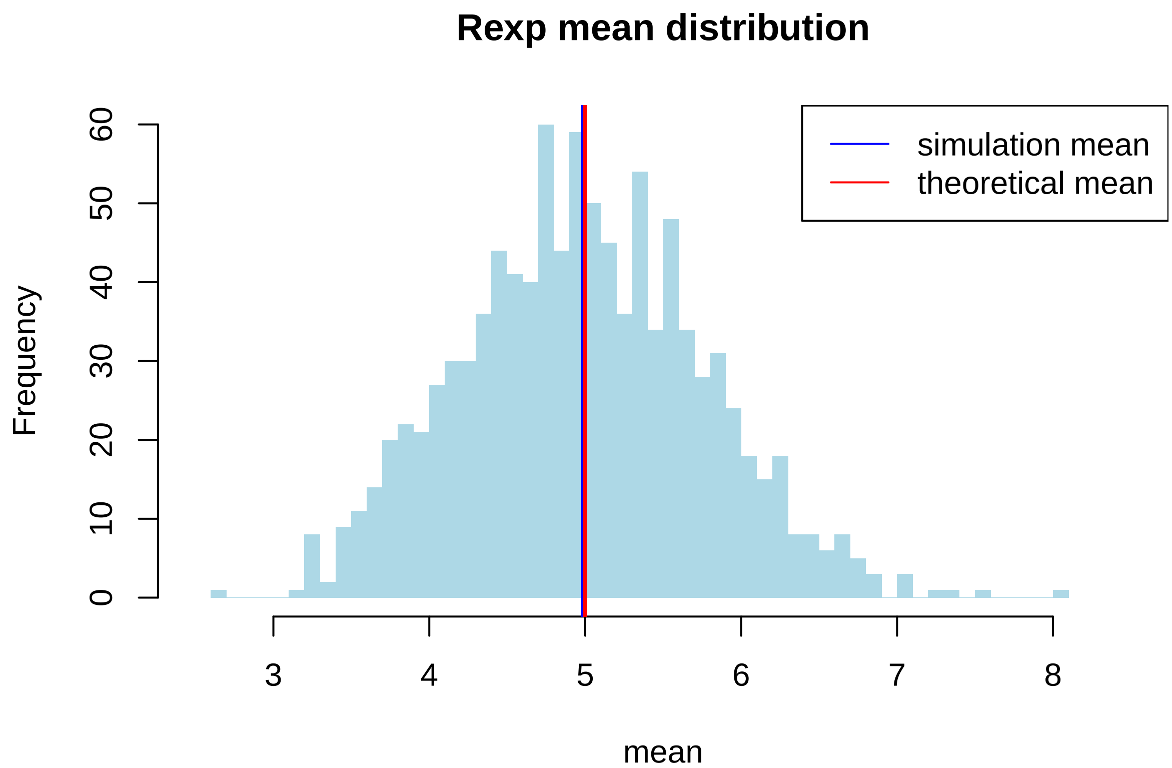
Our sample mean seems very close to the theoretical mean but how to be sure ? Maybe our range of values might help :

```
range(sim_test_Means)
```

```
## [1] 2.688257 8.007416
```

Not bad but we're not totally convinced! Let's create something more visual to see they're close to each other :

```
hist(sim_test_Means , breaks=40 , col="lightblue", border=F , main="Rexp mean distribution", xlab = "mea
abline(v = rexp_sim_mean, lwd="2", col="blue")
abline(v = theoretical_mean, lwd="2", col = "red")
legend('topright', c("simulation mean", "theoretical mean"), lty=c(1,1), col=c("blue", "red"))
```



The histogram confirms what we thought. The two means are very close.

Sample Variance vs Theoretical Variance

We'll start with our sample variance

```
rexp_sim_var <- var(sim_test_Means)
rexp_sim_var
```

```
## [1] 0.6046693
```

Knowing the standard deviation is equals to 1/lambda, it's quite easy to guess the variance formula :

```
theoretical_var <- (1/lambda)^2/n
theoretical_var
```

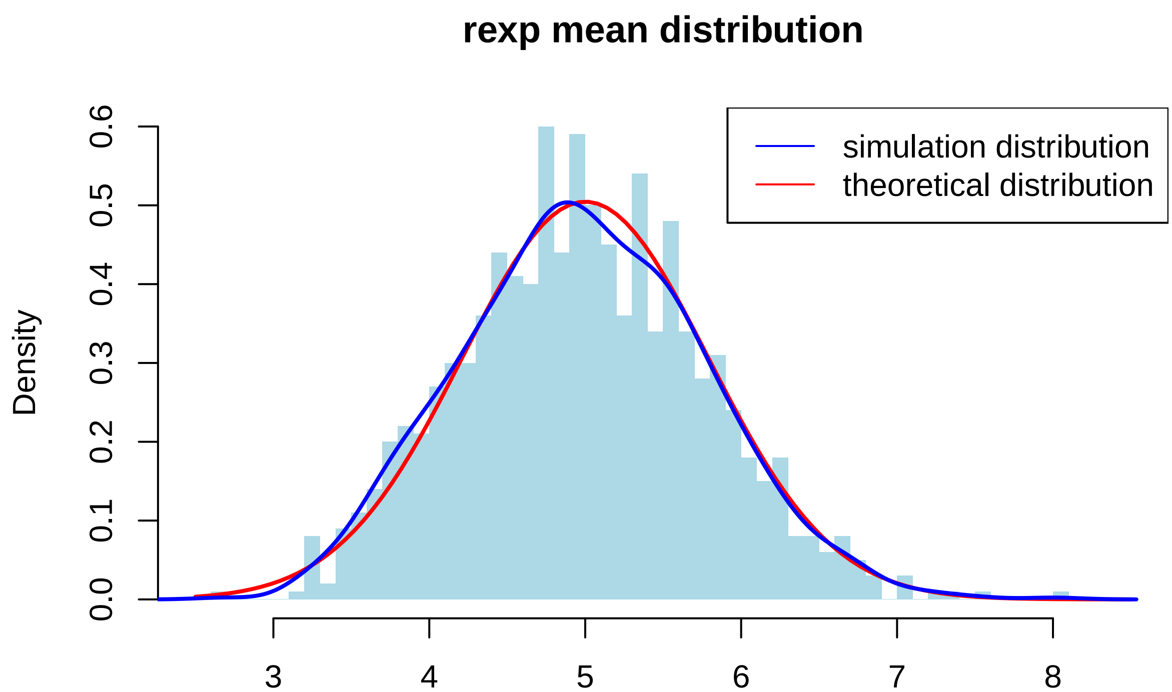
```
## [1] 0.625
```

As for the means, variances are also very close with 0.605 for our sample variance and 6.25 for our theoretical variance.

Is the distribution approximately normal ?

To observe our distribution, we'd to create to be able to compare our sample distribution curve and the theorical distribution on the same plot :

```
hist(sim_test_Means , prob = TRUE, breaks=40 , col="lightblue", border=F , main="rexp mean distribution
curve(dnorm(x, mean=theoretical_mean, sd=sqrt(theoretical_var)), add = TRUE, lwd = 2, col = "red")
lines(density(sim_test_Means), lwd = 2, col = "blue")
legend('topright', c("simulation distribution", "theoretical distribution"), lty=c(1,1), col=c("blue",
```



By watching the plot just above, we can say the plot is approximately normal. However, we can see our sample distribution curve isn't symetric where the theoretical is symetric. We can also add that with a bigger n value, the sample distribution would be even closer to the theoretical.