Statistical Inference Part #1 Simulation Exercise

Valentina Scipione

24 May 2015

Simulations

The exponential distribution is 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 with 1000 simulations.

We begin with 1000 simulated averages of 40 exponentials.

```
# Exponential Distribution
set.seed(3)  # Set seed
n <- 40  # number of exponential random variables
lambda <- 0.2  # lambda for all simulations
nsim <- 1000  # number of simulated averages

mns = NULL
vars = NULL
for (i in 1 : nsim) {
  expd <- rexp(n, lambda)
  mns  <- c(mns, mean(expd))
  vars <- c(vars, var(expd))
}

mean_sample <- mean(mns)  # mean of distribution of averages of 40 exponentials
mean_theoretical <- 1/lambda  # mean from analytical expression
```

```
summary(mns)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   2.567   4.406   4.945   4.987   5.522   7.457
```

```
summary(vars)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.629  17.390  22.280  24.730  30.060  81.720
```

```
mean_sample
```

```
## [1] 4.98662
```

```
mean_theoretical
```

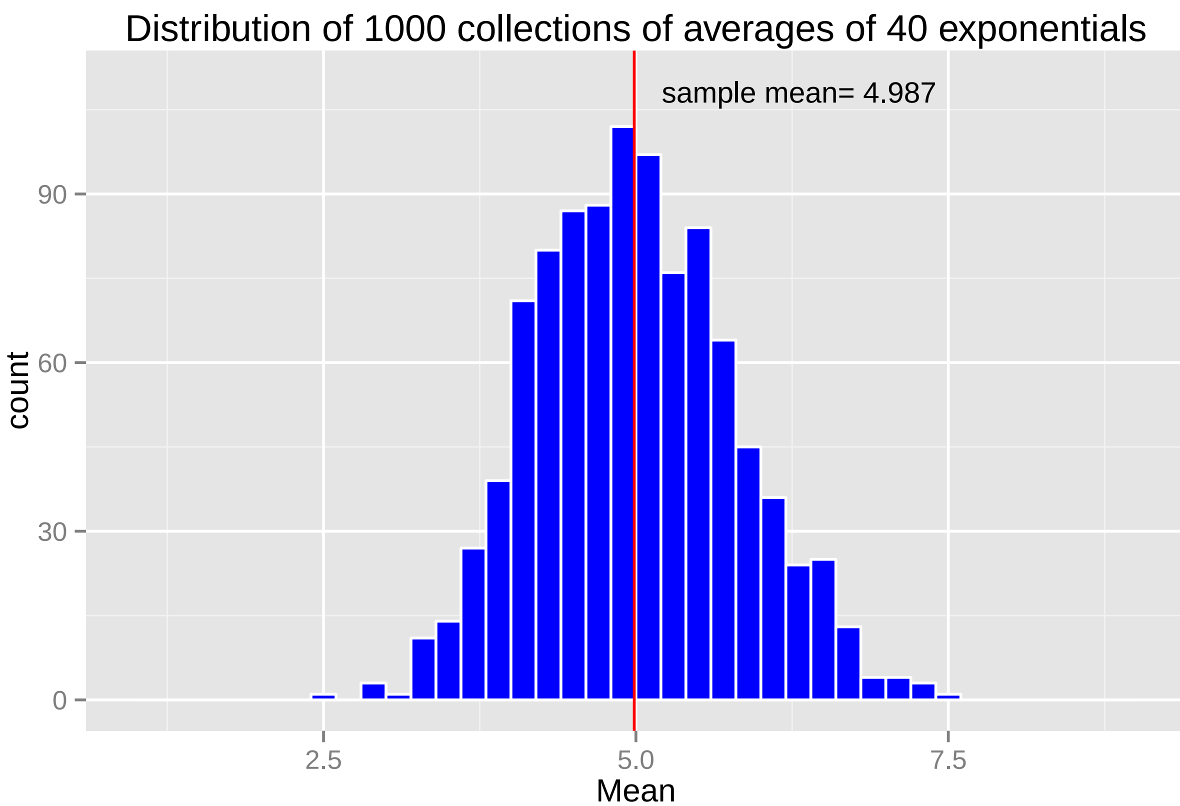
```
## [1] 5
```

Results

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

Plot of the Distribution of 1000 collections of averages of 40 exponentials

```
# Plot the distribution of 1000 collections of averages of 40 exponentials
library(ggplot2)
g1 <- qplot(mns,  fill=I("blue"), color=I("white"), geom="histogram", xlab="Mean", binwidth=0.2, xlim=c
            main="Distribution of 1000 collections of averages of 40 exponentials")
g1 <- g1 + geom_vline(xintercept = mean_sample, color="red")
g1 <- g1 + geom_text(mapping=aes(x=mean_sample, y=110, label=paste("sample mean=",round(mean_sample,3))
g1
```



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

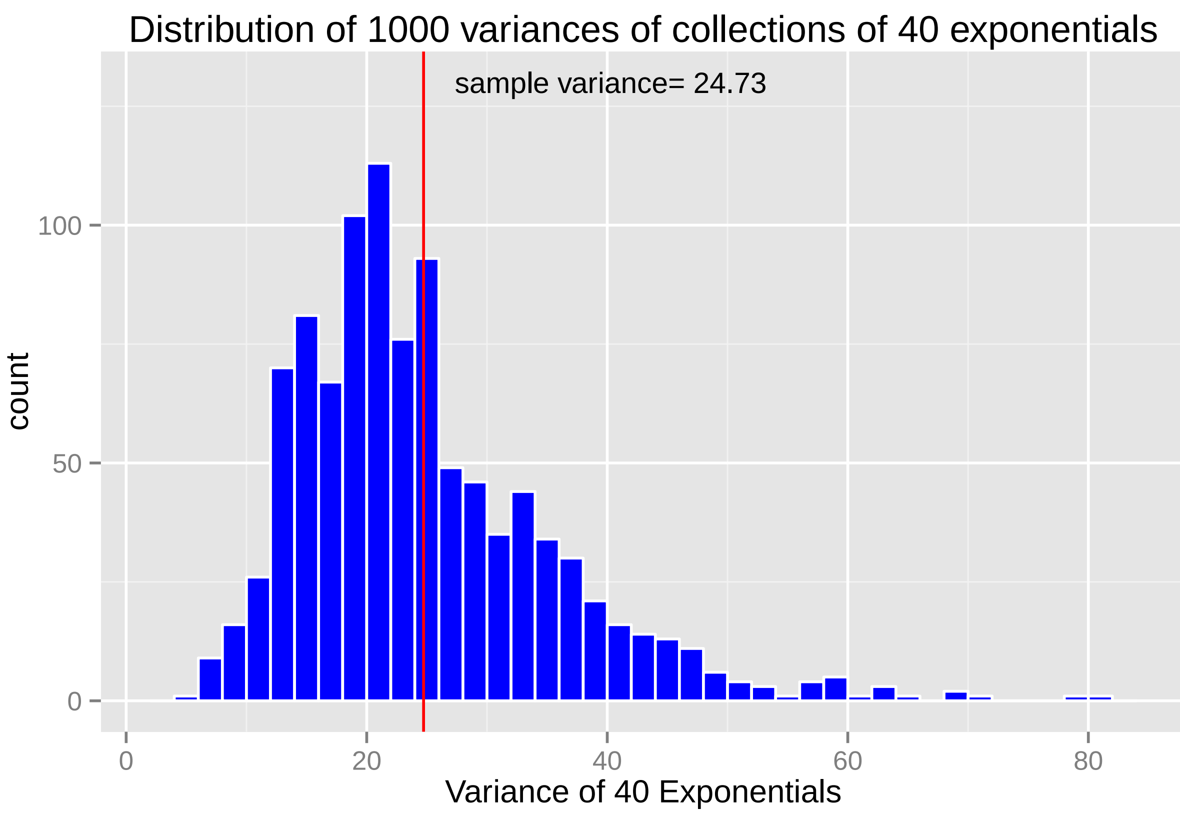
```
# Plot the distribution of variances
vars_sample <- mean(vars)
vars_theoretical <- (1/lambda)^2 #variance= std^2
vars_sample
```

```
## [1] 24.72853
```

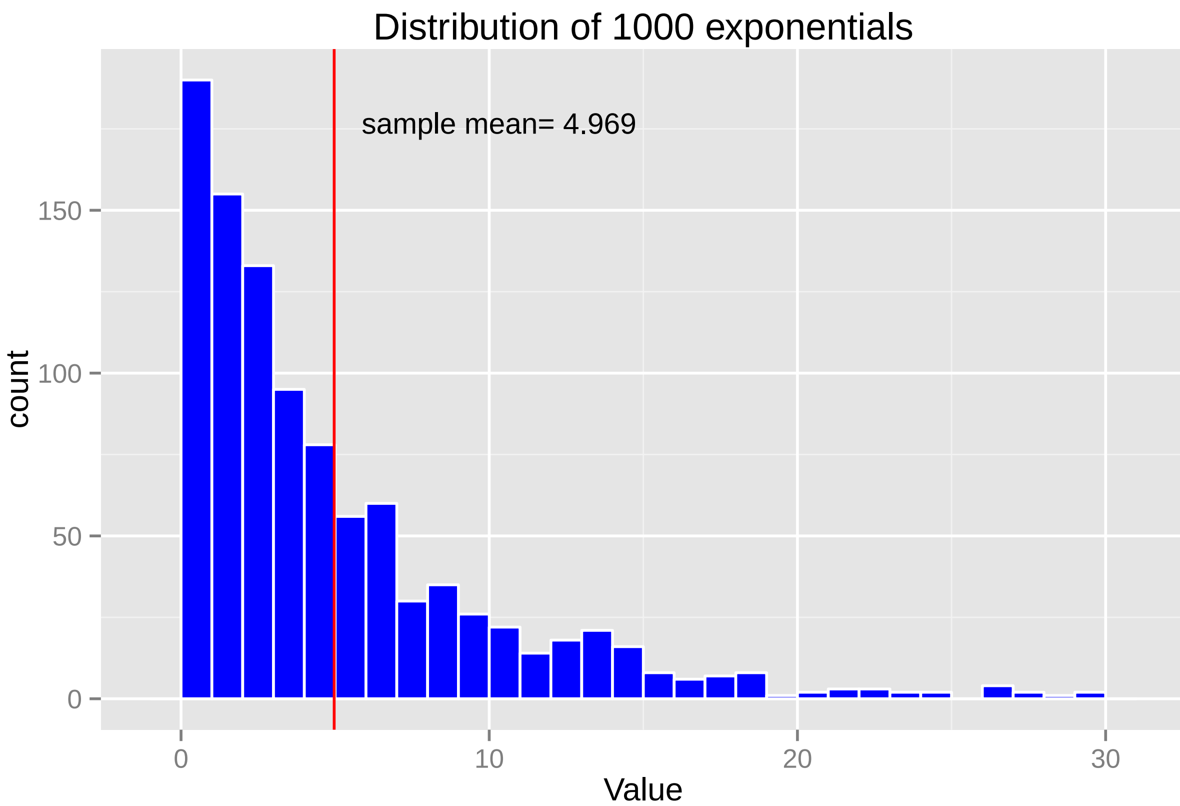
```
vars_theoretical
```

```
## [1] 25
```

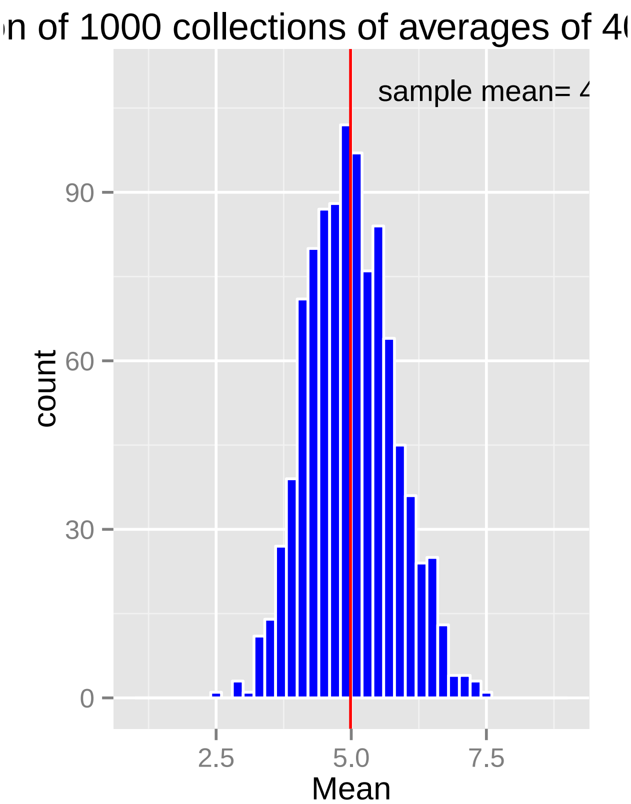
```
g2 <- qplot(vars,  fill=I("blue"), color=I("white"), geom="histogram", binwidth=2, xlab="Variance of 40
            main="Distribution of 1000 variances of collections of 40 exponentials")
g2 <- g2 + geom_vline(xintercept = vars_sample, color="red")
g2 <- g2 + geom_text(mapping=aes(x=vars_sample, y=130, label=paste("sample variance=",round(vars_sample
g2
```

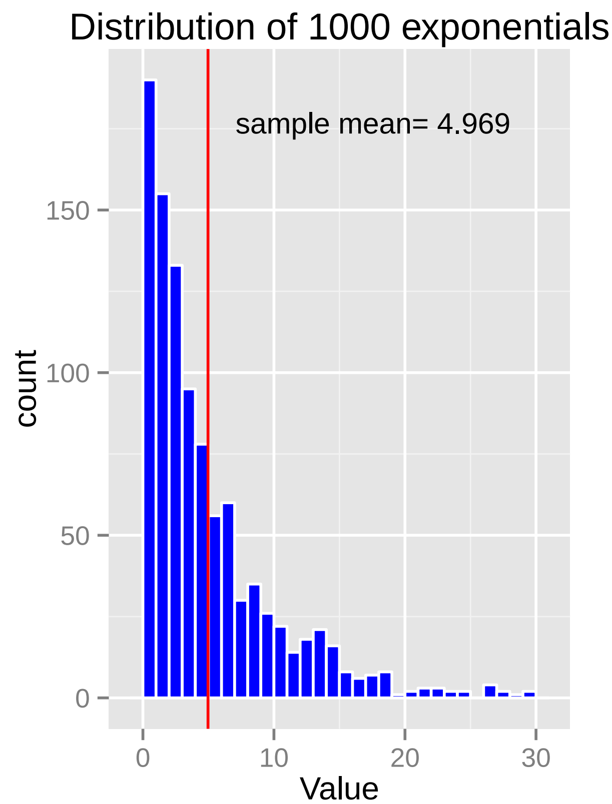



```
# Compare with the distribution of 1000 exponentials
expd <- rexp(nsim, lambda)
expd_mean = mean(expd)
expd_var = var(expd)
g3 <- qplot(expd,  fill=I("blue"), color=I("white"), geom="histogram", xlab="Value", binwidth=1,
            main="Distribution of 1000 exponentials")
g3 <- g3 + geom_vline(xintercept = expd_mean, color="red")
g3 <- g3 + geom_text(mapping=aes(x=expd_mean, y=180, label=paste("sample mean=",round(expd_mean,3))), s
g3
```




```r
library(grid)
library(gridExtra)
grid.arrange(g1, g3, ncol=2)
```

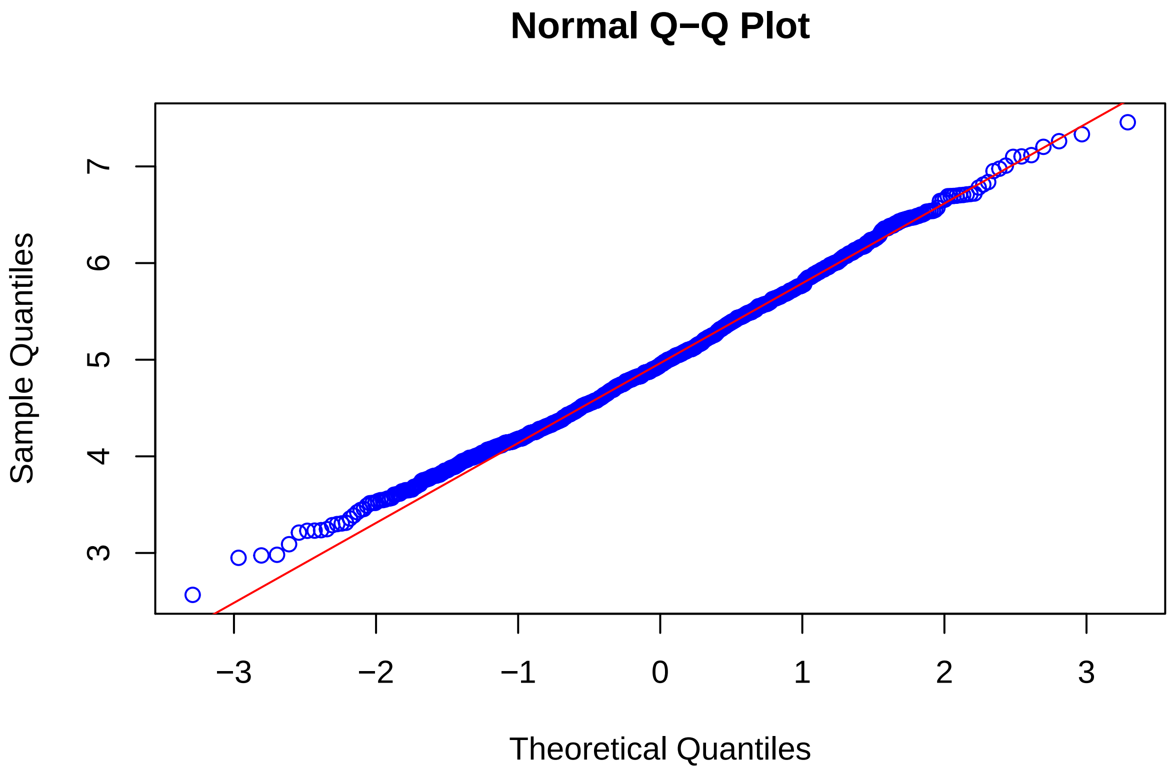




"'

3. Show that the distribution is approximately normal

```
# use qqplot and qqline to compare the distribution of averages of 40 exponentials to a normal distribu
qqnorm(mns, col="blue")
qqline(mns, col = 2)
```


Normal Q−Q Plot
Sample Quantiles
3
4
5
6
7
−3
−2
−1
0
1
2
3
Theoretical Quantiles