Simulation from an exponential distribution

Rudy Veenhoff

Overview

In this paper we will investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Simulations

The exponential distribution can be simulated in R with rexp(n, λ), where λ is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. We set $\lambda = 0.2$ for all of the simulations. We will investigate the distribution of averages of 40 exponentials. Note that we will do a thousand simulations.

```
set.seed(11235813)
noSim <- 1000
lambda <- 0.2
sampleSize <- 40

## the vector of means of the 1000 simulations
means <- NULL
for(i in 1:noSim) means <- c(means,mean(rexp(sampleSize,lambda)))
```

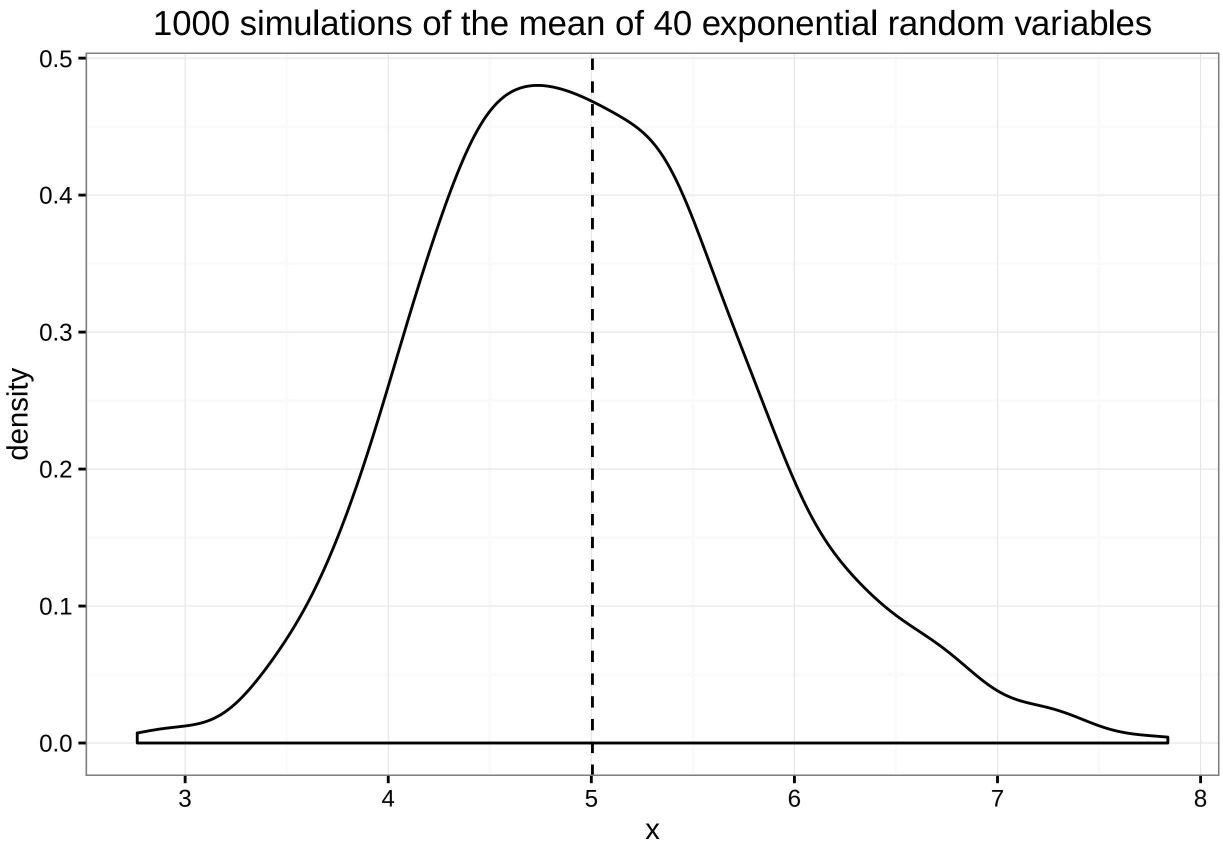
Sample Mean versus Theoretical Mean

Let $X_1, X_2, \ldots, X_{40}$ be 40 *iid* exponential random variables with rate $\lambda = 0.2$ (so $E(X_i) = 1/\lambda$ and $Var(X_i) = 1/\lambda^2$ for $i = 1, 2, \ldots, 40$). The Central Limit Theorem tells us that the mean of these 40 variables is approximately normal:

$$\bar{X}_{40} \sim N(1/\lambda, 1/(40\lambda^2)).$$

Plugging in $\lambda = 0.2$ results in $\bar{X}_{40} \sim N(5, 0.625)$. Now that we know the distribution of the sample mean, let's take a look at plot of the simulation.

```
## plotting the density
library(ggplot2)
g <- ggplot(data.frame(means=means),aes(x=means))
g + theme_bw(base_size = 11, base_family = "") +
  geom_density() +
  geom_vline(aes(xintercept=mean(means)),linetype="dashed") +
  labs(title="1000 simulations of the mean of 40 exponential random variables") +
  labs(x="x")
```



The dashed line indicates the simulated mean, i.e. 5.0054629. Note this is very close to the theoretical mean of 5.

Sample Variance versus Theoretical Variance

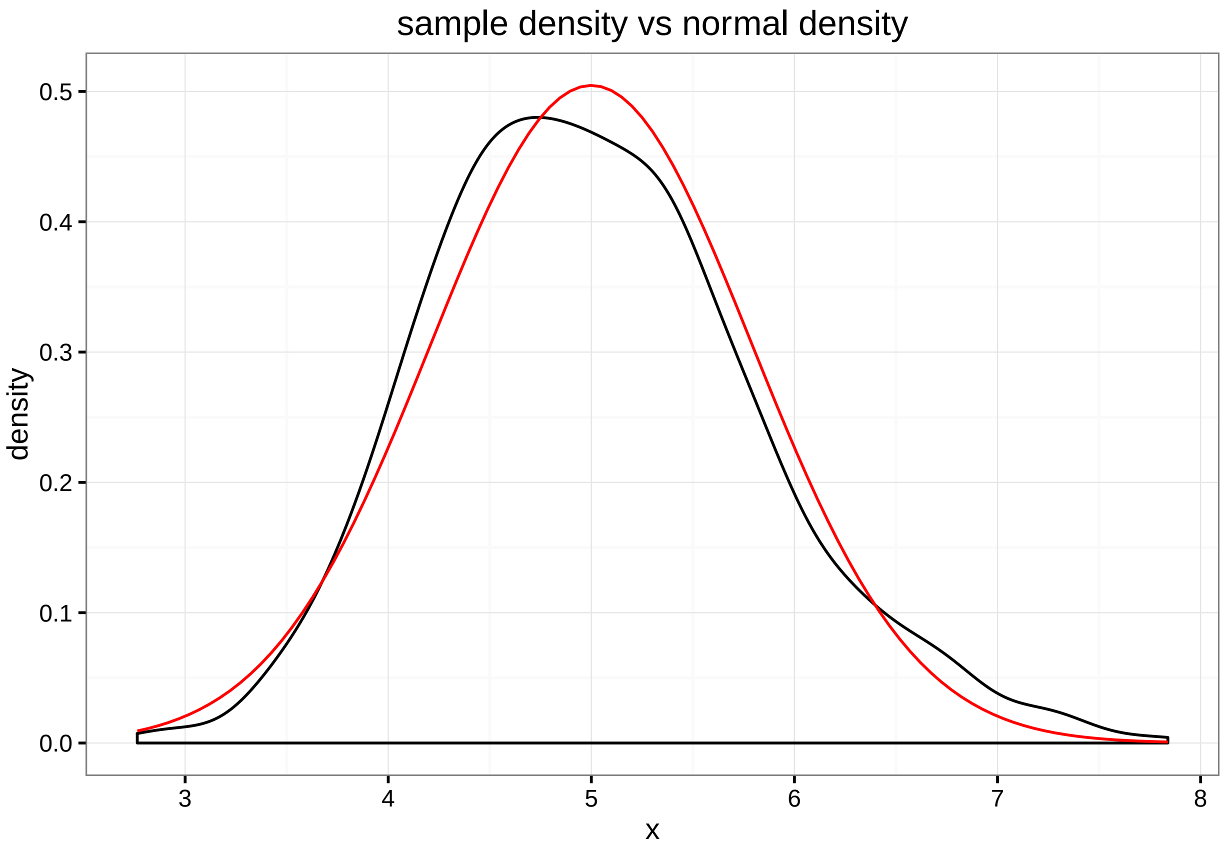
The sample variance is:

```
var(means)
```

```
## [1] 0.648105
```

Again, this is pretty close to our theoretical variance of 0.625. We compare the plots of the theoretical and the sample distribution.

```
g + theme_bw(base_size = 11, base_family = "") +
  geom_density() +
  labs(title="sample density vs normal density") +
  labs(x="x") +
  stat_function(fun = dnorm, args = list(mean = 5, sd = sqrt(0.625)), colour="red")
```

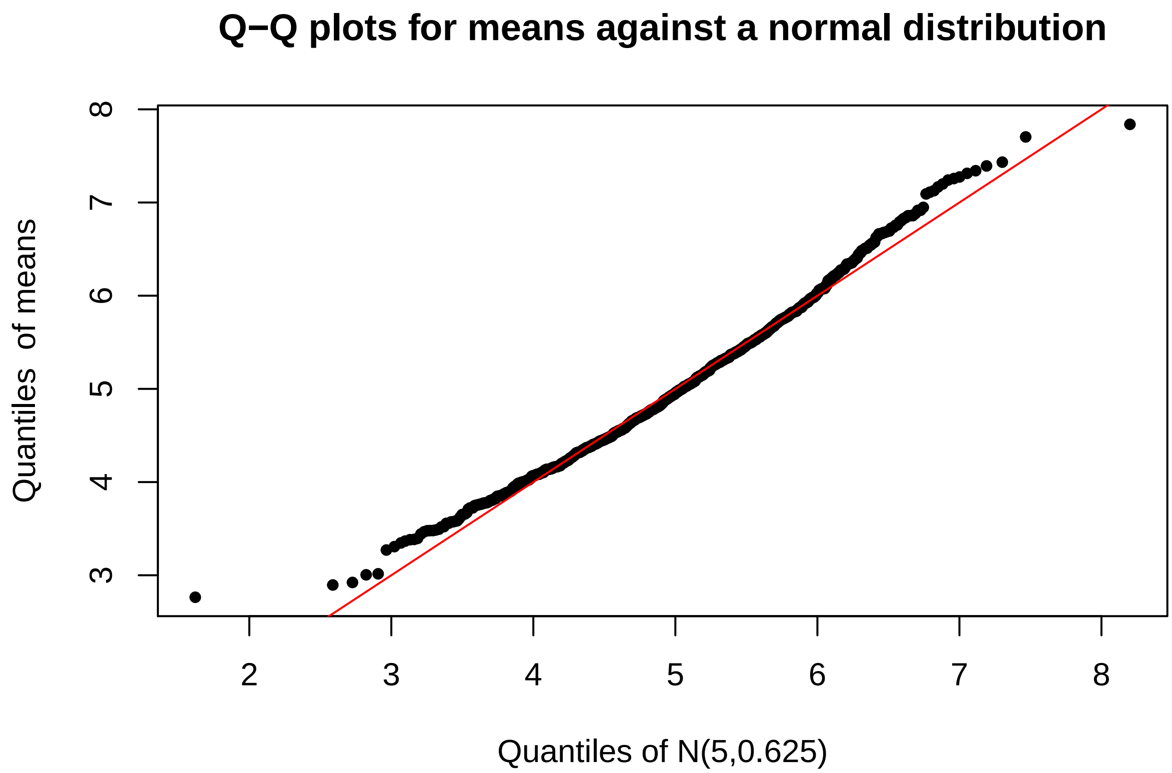


Distribution

In this section we compare the normal distribution to our simulated distribution. Theoretically they should be approximately the same. We've already seen in the previous section that the density of the simulated data looks very much like a normal density. In this section we investigate the Q-Q plot of means versus a sample of 100000 random normals with mean 5 and standard deviation $\sqrt{0.625}$.

```
qqplot(rnorm(100000, mean=5,sd=sqrt(0.625)), means, pch=20,
       xlab = "Quantiles of N(5,0.625)",
       ylab = "Quantiles  of means",
       main = "Q-Q plots for means against a normal distribution")

## adds the line y=x.
abline(a=0,b=1,col="red",lwd=1)
```



We note that the Q-Q plot and the (red) line $y = x$ are almost overlapping. This is gives us strong reasons to believe our theoretical deduction that the mean is approximately normal.