Statistical Inference - Assignemnt

Daniel Bader

Fri Jun 19 00:03:02 2015

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

Exponential Distribution

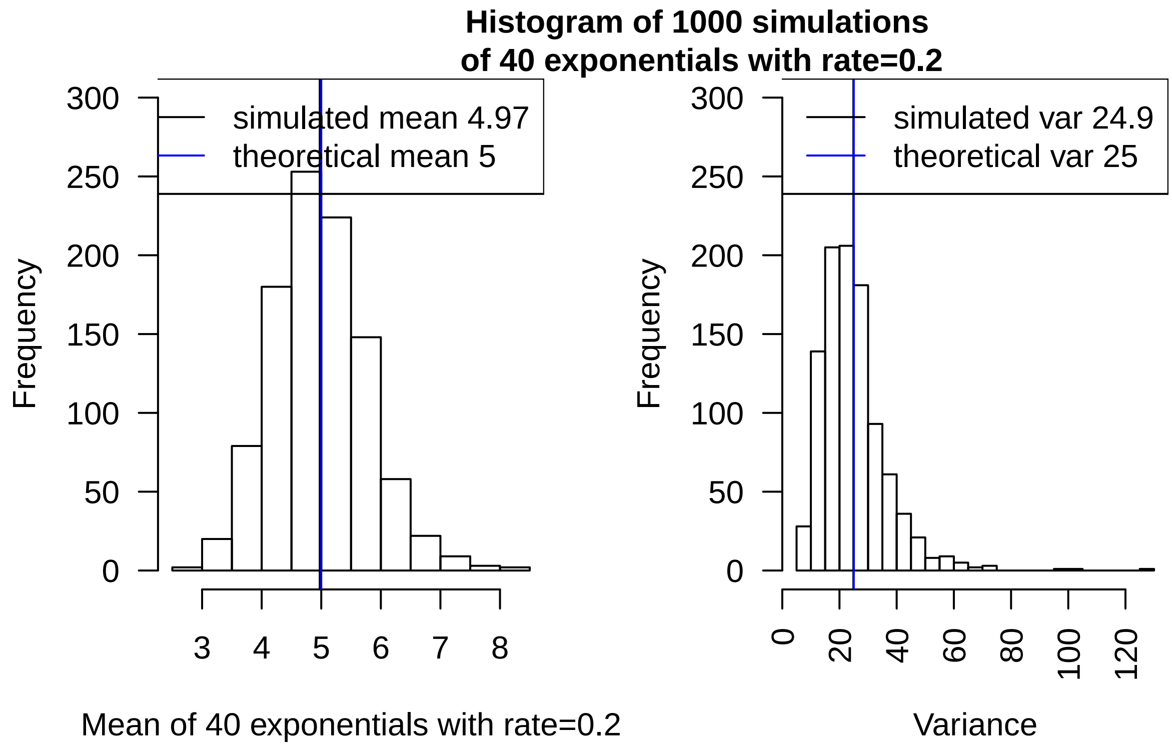
Define basic variables for the simulation:

```
suppressPackageStartupMessages(library(geneplotter))
LAMBDA=0.2
nummeans=40
nsimul=1000
```

Mean and variance: Sample versus Theory

```
exp_means = exp_vars = NULL
# simul data
set.seed(42)
for (i in 1 : nsimul){
    exp_means = c(exp_means, mean(rexp(nummeans, rate=LAMBDA)))
    exp_vars = c(exp_vars, var(rexp(nummeans, rate=LAMBDA)))
}
limits=c(0,300)
par(mfrow=c(1,2))
# mean
hist(exp_means,
     xlab=paste0('Mean of ',nummeans,' exponentials with rate=',LAMBDA),
     main='', ylim=limits, las=1
     )
vvalues= c(mean(exp_means),1/LAMBDA)
vcols= c('black','blue')
abline(v= vvalues, col=vcols)
legend('topright', paste(c('simulated mean','theoretical mean'), signif(vvalues,3)),
       col=vcols, lwd=1)
# var
hist(exp_vars,
     xlab=paste0('Variance'),
     main='', breaks=30, ylim=limits, las=2
)
vvalues= c(mean(exp_vars), 1/LAMBDA^2)
abline(v= vvalues, col=vcols)
legend('topright', paste(c('simulated var','theoretical var'), signif(vvalues,3)),
```

```
        col=vcols, lwd=1
)
par(mfrow=c(1,1))
mtext(paste0('Histogram of ',nsimul,' simulations \nof ',
             nummeans,' exponentials with rate=',LAMBDA), cex=1, font=2)
```

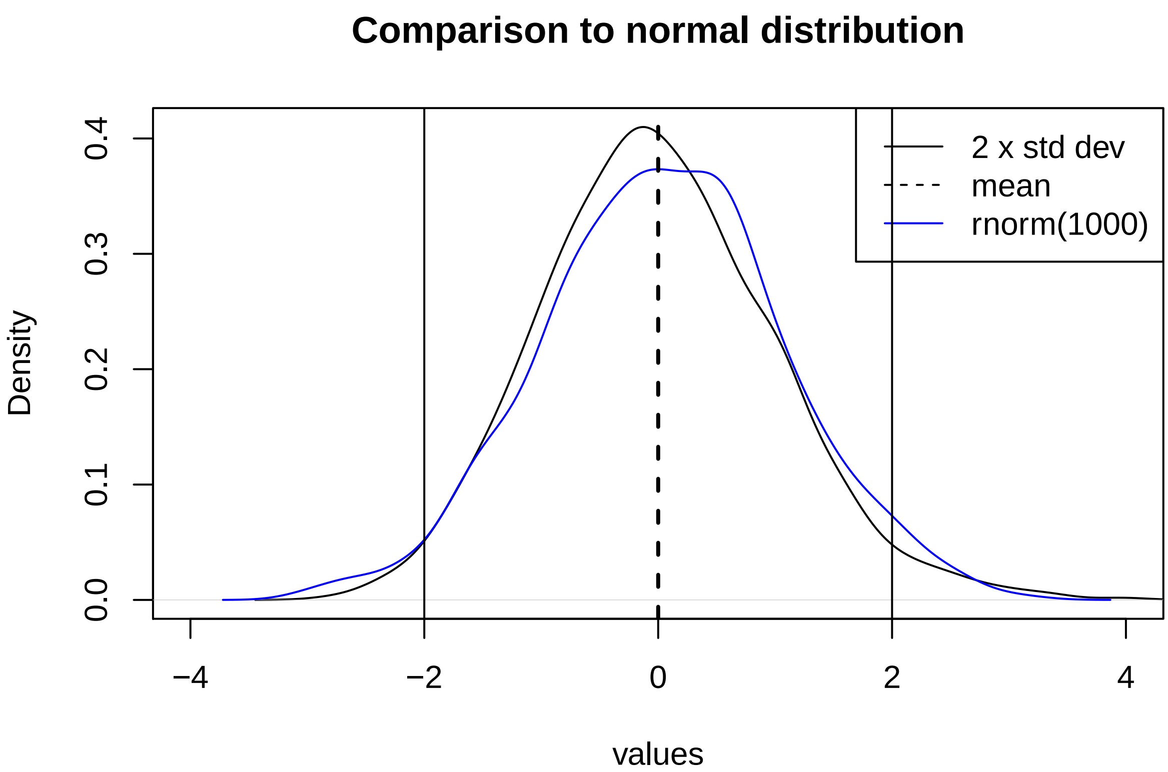


The above code is computing 1000 times the mean and variance of 40 random draws from an exponential distribution with rate=0.2. The simulated mean is very close to the theoretical mean=1/rate. The simulated mean of variances is very close to the theoretical var=1/rate^2. This leads to the conclusion that 1000 simulations are enough to see the Gaussian distribution predicted by the central limit theorem.

Approximately Normal Distribution

```
quasi_normal= (exp_means - mean(exp_means))/sd(exp_means)
multidensity( list(quasi_normal, rnorm(nsimul)),
     main='Comparison to normal distribution',
     xlab='values',
     xlim=c(-4,4),
     col=c('black', 'blue'),
     legend= list(x='topright',
                  legend=c('2 x std dev', 'mean', 'rnorm(1000)'),
                  lty=c(1,2,1),
                  col=c('black', 'black', 'blue') )
     )
```

```r
tmplines= c(1,1,2)
abline(v=c(c(-2,2)*sd(quasi_normal), mean(quasi_normal)), lwd=tmplines, lty=tmplines)
```



The above code transforms the distribution of means (X) for 40 exponentials into a standard normal by the following equation:

$$\frac{X - \mu}{SE}$$

where μ and **SE** are the mean and standard deviation of the simulated distribution, respectively.