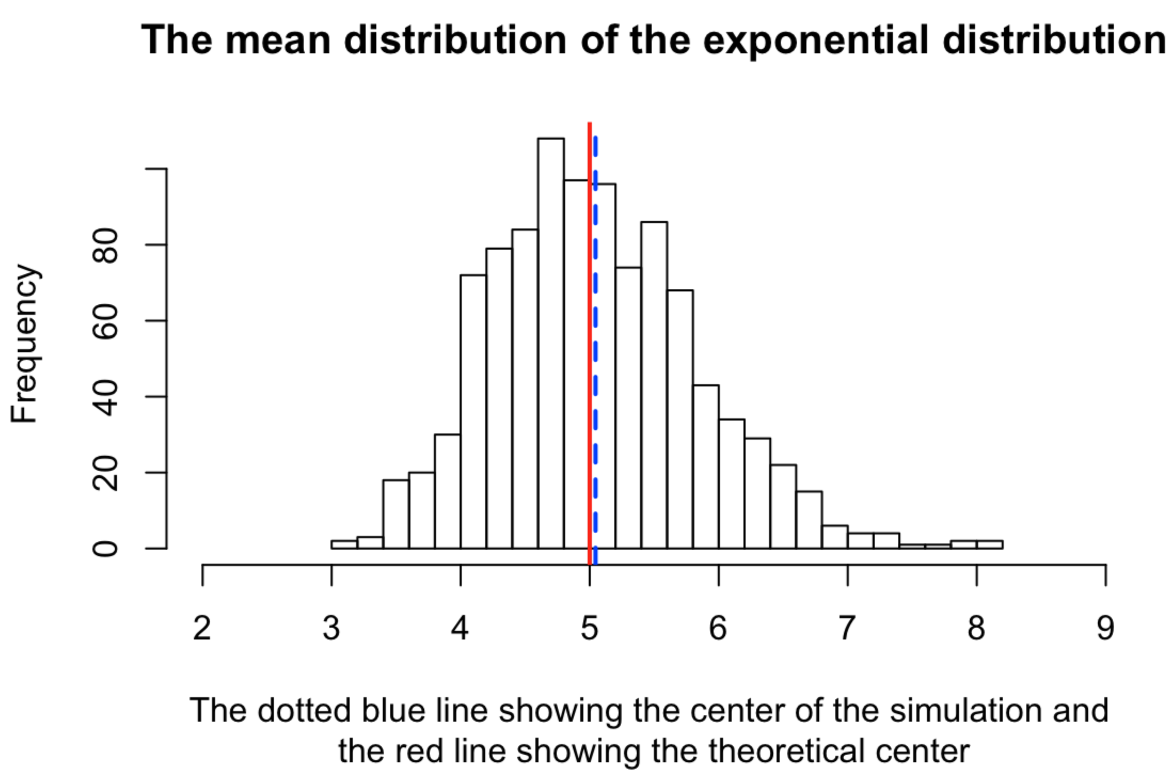
Statistical Inference Project

Instruction: 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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

1. Show where the distribution is centered at (*see the dotted blue line*) and compare it to the theoretical center of the distribution (*see the red line*).

```
nosim <- c(1000,1000, 1000); size <- c(10, 40, 80); n<- size[2]; lambda <-0.2; set.seed(10); library(ggplot2)
data <- sapply(1:nosim[2], function(m) {xvals <- rexp(n, lambda); return(c(mean(xvals),sd(xvals), 100))})
data <-data.frame(mean=data[1,], sd=data[2,], lowlimit=data[1,]-1.96*data[2,]/sqrt(n), uplimit=data[1,]+1.96*data[2,]/sqrt(n))
hist(data$mean, breaks=25, xlim=c(2,9),main="The mean distribution of the exponential distribution", xlab="", sub="The dotted blue line showing the center of the simulation and \nthe red line showing the theoretical center");
abline(v=mean(data$mean), lwd=2,col="blue", lty=2)  #sampling distribution center
abline(v=1/lambda, lwd=2, col="red") #Theoretical center of the distribution
```



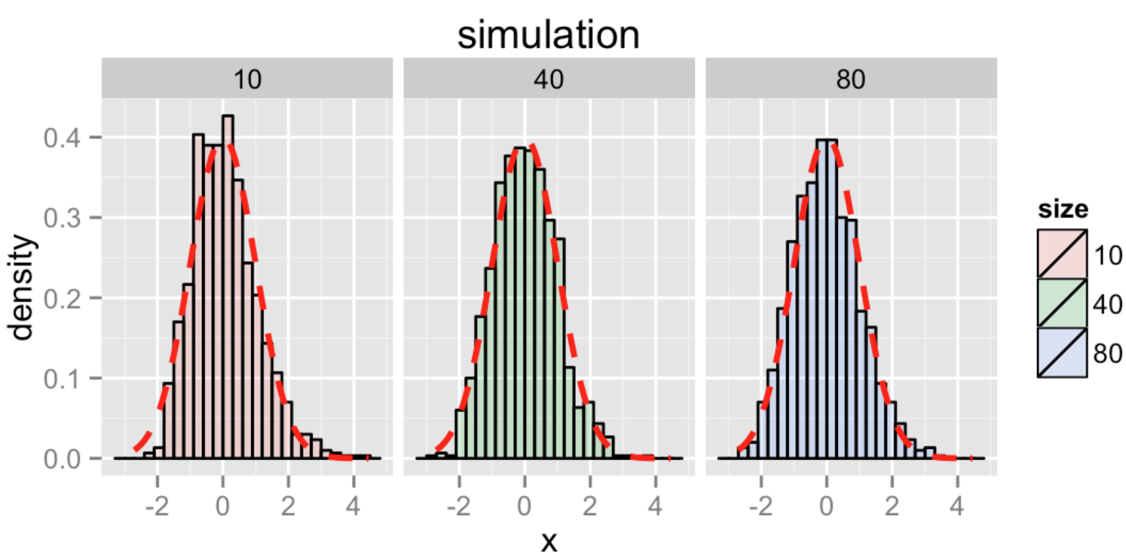
2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
data.frame(sampling.distribution.variance=var(data$mean),theoretical.mean.variance=(1/(lambda^2
*n))) # compare the variance of the sampling distribution and the theoretical variance
```

```
##   sampling.distribution.variance theoretical.mean.variance
## 1                         0.6373                     0.625
```

3. Show that the distribution is approximately normal (see the red dotted line below) when the sample size is increased (>30).

```
cfunc <- function(x, n) {sqrt(n)*(mean(x)-1/lambda)*lambda}
dat <- data.frame(
      x = c(apply(matrix(rexp(nosim[1]*size[1],lambda), nosim[1]), 1, cfunc, size[1]),
            apply(matrix(rexp(nosim[2]*size[2],lambda), nosim[2]), 1, cfunc, size[2]),
            apply(matrix(rexp(nosim[3]*size[3],lambda), nosim[3]), 1, cfunc, size[3])),
   size = factor(sapply(nosim, function(m) rep(size, rep(m,3)))))
g <- ggplot(dat, aes(x = x, fill = size)) + geom_histogram(alpha = .20, binwidth=.3, colour = "
black", aes(y = ..density..)) +
     stat_function(fun = dnorm, size = 1, lty=2, col = "red") + facet_grid(. ~ size) ; g + ggti
tle("simulation")
```



4. Evaluate the coverage of the confidence interval for 1/lambda:

```
sprintf("The coverage of the confidence interval for 1/lambda: %f",  sum(data$lowlimit<=1/lambd
a & data$uplimit>=5)/nosim[2])
```

```
## [1] "The coverage of the confidence interval for 1/lambda: 0.941000"
```