Statistical_Inference_Project01

Part#1 Simulation Exercise Simulations The exponential distribution can be simulated in R with rexp(n, lambda) where lambda(λ) is the rate parameter. The mean of exponential distribution is 1/λ and the standard deviation is also 1/λ. For this simulation, we set λ = 0.2. In this simulation, we investigate the distribution of averages of 40 exponential(0.2)s. Lets start by doing a thousand simulated averages of 40 exponentials

```
# Set seed
set.seed(3)
lambda <- 0.2
# We perform 1000 simulations with 40 samples
sample_size <- 40
simulations <- 1000
# Lets do 1000 simulations
simulated_exponentials <- matrix(rexp(simulations*sample_size, rate=lambda), simulations, sample_s
ize)
# Averages of 40 exponentials
row_means <- rowMeans(simulated_exponentials)
```

Results

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

```
# mean of distribution of averages of 40 exponentials
mean(row_means)
```

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

```
# mean from analytical expression
1/lambda
```

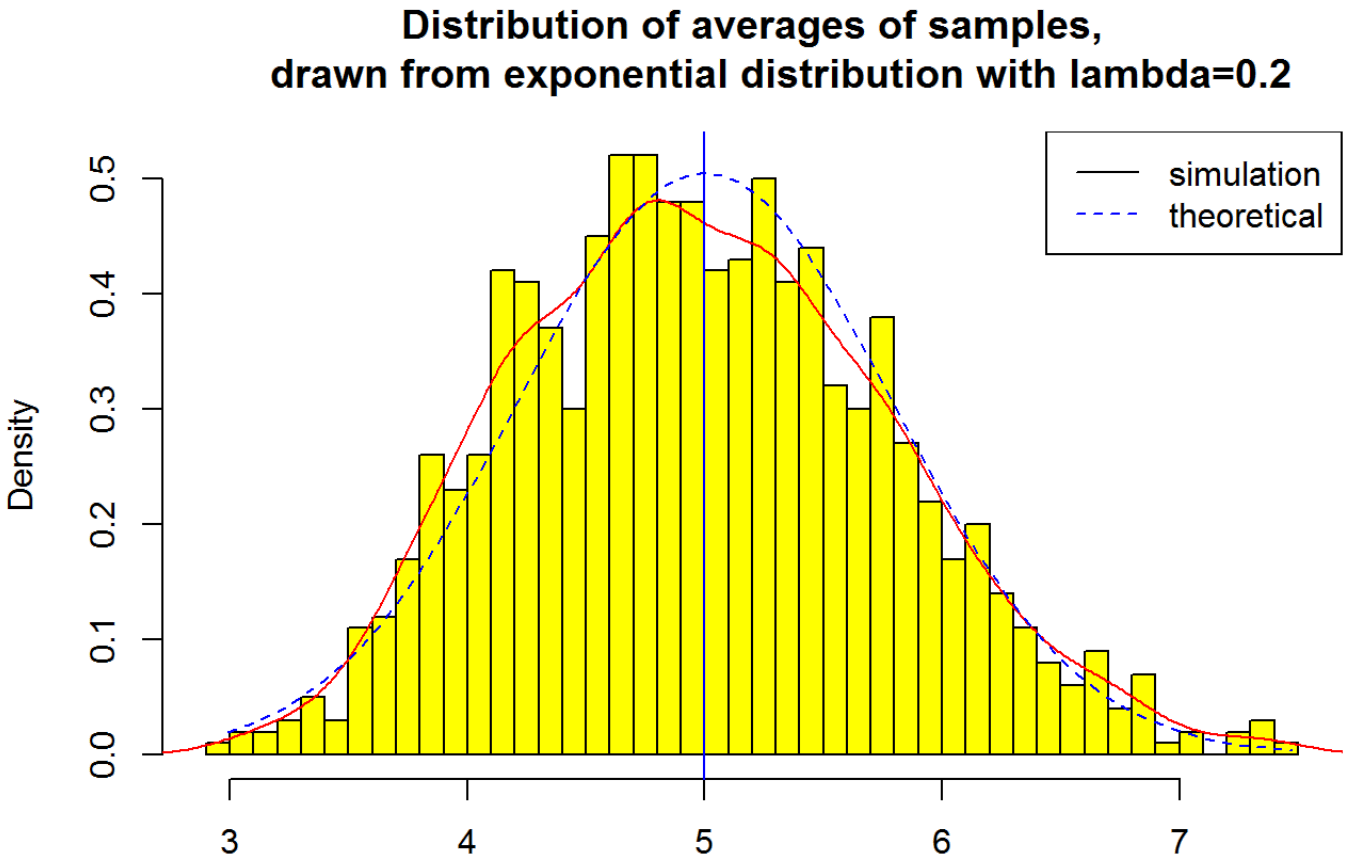
```
## [1] 5
```

The distribution of sample means is shown below:

```
# plot the histogram of averages
hist(row_means, breaks=50, prob=TRUE,
     main="Distribution of averages of samples,
     drawn from exponential distribution with lambda=0.2",
     xlab="", col="yellow")
# density of the averages of samples
lines(density(row_means), col="red")

# theoretical center of distribution
abline(v=1/lambda, col="blue")
# theoretical density of the averages of samples
xfit <- seq(min(row_means), max(row_means), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample_size)))
lines(xfit, yfit, pch=22, col="blue", lty=2)

# add legend
legend('topright', c("simulation", "theoretical"), lty=c(1,2), col=c("black", "blue"))
```



Therefore, the distribution of averages of 40 exponentials is centered at 4.9866 and the same is close to the theoretical center of the distribution, which is λ???1 = 5. **2. Show how variable it is and compare it to the theoretical variance of the distribution**

```
# standard deviation of distribution of averages of 40 exponentials
sd(row_means)
```

```
## [1] 0.7910484
```

```
# standard deviation from analytical expression
(1/lambda)/sqrt(sample_size)
```

```
## [1] 0.7905694
```

```
# Variance of the sample mean
var(row_means)
```

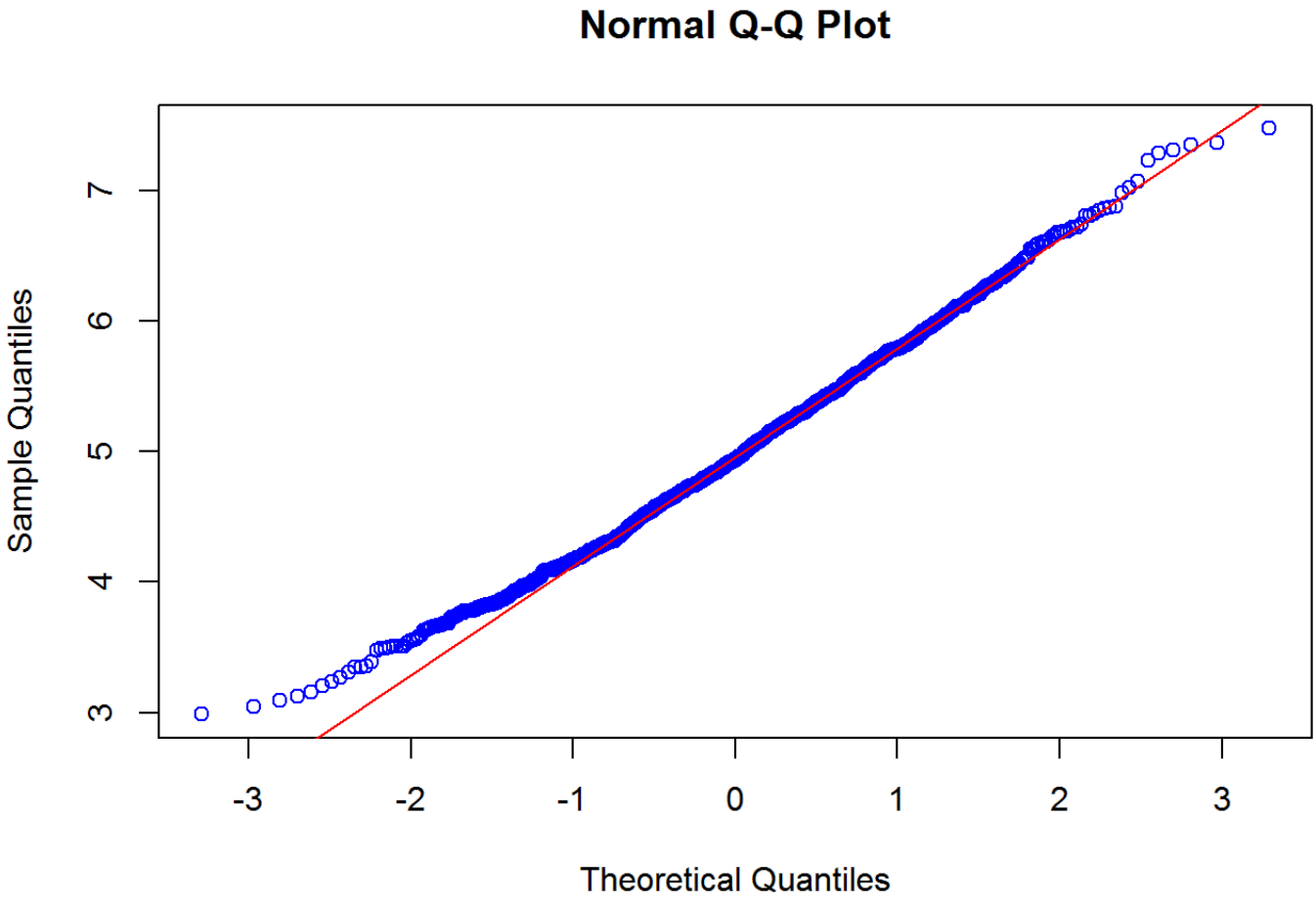
```
## [1] 0.6257575
```

```
# Theoritcal variance of the distribution
1/((0.2*0.2) * 40)
```

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

Therefore, the variability in distribution of averages of 40 exponentials is close to the theoretical variance of the distribution. The variance of sample means is 0.6258 where as the theoretical variance of the distribution is σ2/n =1/(λ 2n) = 1/(0.04 × 40) = 0.625. **3. Show that the distribution is approximately normal.**

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



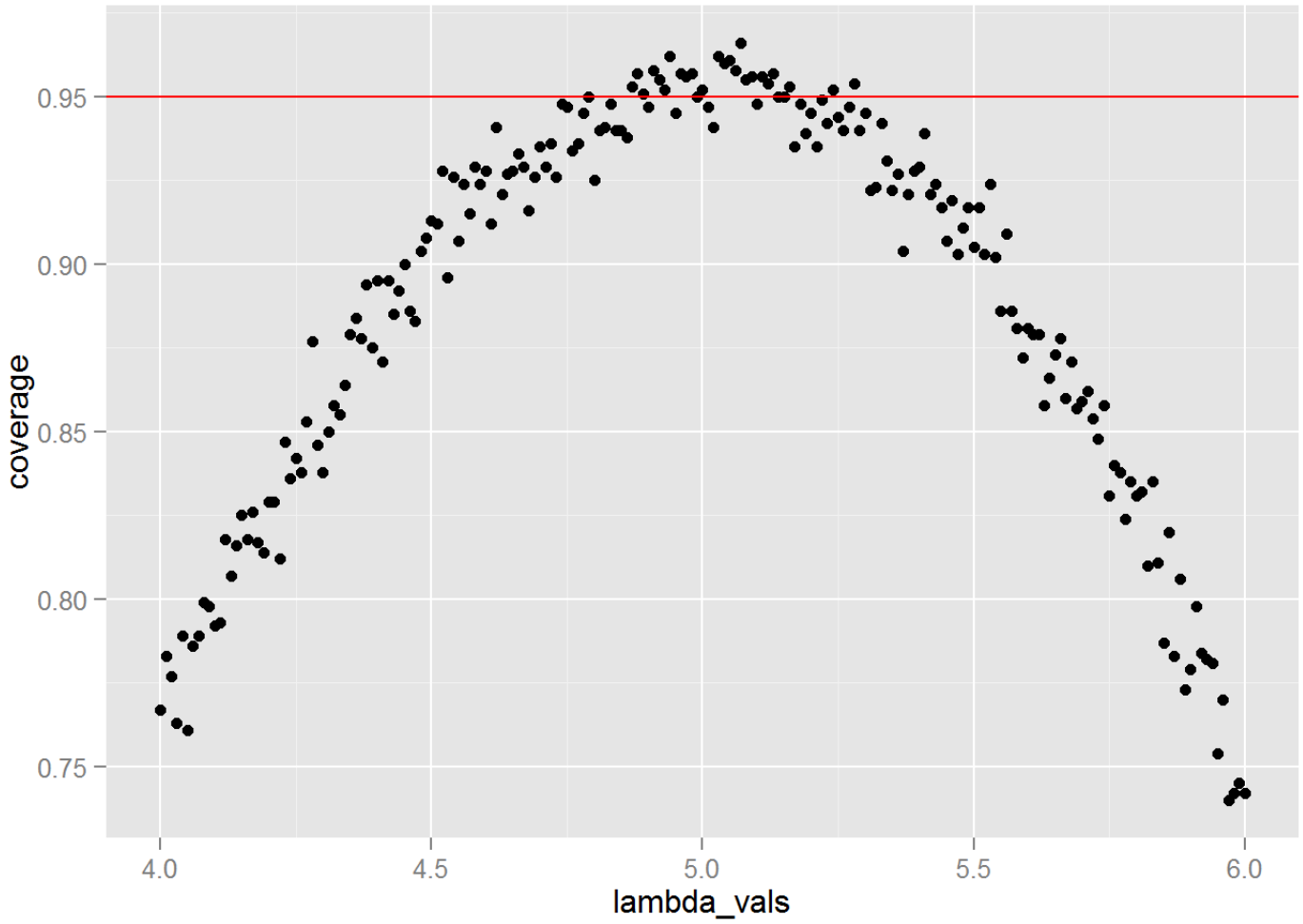
Due to the central limit theorem, the averages of samples follow normal distribution. The figure above also shows the density computed using the histogram and the normal density plotted with theoretical mean and variance values. Also, the q-q plot suggests the distribution of averages of 40 exponentials is very close to a normal distribution. **4. Evaluate the coverage of the confidence interval**

```
# Evaluate coverage of the confidence interval
lambda_vals <- seq(4, 6, by=0.01)
coverage <- sapply(lambda_vals, function(lamb) {
    mu_hats <- rowMeans(matrix(rexp(sample_size*simulations, rate=0.2),
                               simulations, sample_size))
    ll <- mu_hats - qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    ul <- mu_hats + qnorm(0.975) * sqrt(1/lambda**2/sample_size)
    mean(ll < lamb & ul > lamb)
})

# Plot coverage
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.1.2
```

```
qplot(lambda_vals, coverage) + geom_hline(yintercept=0.95, col=2)
```


0.95
0.90
0.85
0.80
0.75
coverage
4.0
4.5
5.0
5.5
6.0
lambda_vals