Course Report of Exponential Distribution

Xudong Zhang

September 9, 2014

Theoretical Statistics of Exponential Distribution

```
n <- 40
lambda <- 0.2
theory_mean <- 1 / lambda
theory_std_dev <- 1 / lambda
theory_std_err <- theory_std_dev / sqrt(n)
```

First, let's look at the theoretical statistical value of the exponential distribution. When lambda is set as 0.2, the theoretical mean is 5 and the standard deviation is also 5. As far as the distribution of averages of 40 exponential(0.2)s is concerned, the mean is 5 and the standard error is 0.7906.

Distribution of Sample Means of Exponential Distribution

Next we sampled the means of exponential distribution by 1000 times. The average is from 40 exponential(0.2)s. We use the ggplot2 library to display our results.

```
library(ggplot2)
set.seed(1014093)
nosim <- 1000
n <- 40
samples <- apply(matrix(rexp(nosim * n, rate = lambda), nosim), 1, mean)
```

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

```
samples_mean <- mean(samples)
samples_sd <- sd(samples)
mean_diff <- samples_mean - theory_mean
```

The mean of average of simulation is 5.0182, which is 0.0182 larger than the theoretical center of the distribution.

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

```
var_diff <- samples_sd^2 - theory_std_err^2
std_err_diff <- samples_sd - theory_std_err
```

The variance of average of simulation is 0.6128, which is -0.0122 from the theoretical variance of the distribution. Therefore both the mean and variance between sample distribution and theoretical distribution are very close.

3. Show that the distribution is approximately normal.

Figure 1 shows the distribution of the simulated means of exponential distribution(blue curve), and the normal distribution (green curve). we found that the distribution of the simulation is approximately normal

To further test the normality of the distribution of the means, we use approaches: (1) Shapiro-Wilk test, (2) Anderson-Darling test and (3) Q-Q plot.

```
library(nortest)
shapiro.test(samples)$p.value
```

```
## [1] 0.0003287
```

```
ad.test(samples)$p.value
```

```
## [1] 0.0002988
```

The p-value in both Shapiro-Wilk test and Anderson-Darling test is less than 5%. Since the null hypothesis this that the distribution is normally distributed, we reject the null hypothesis, and therefore the simluation does not *exactly* have a normal distribution. However, from the Q-Q plot we can still found the simulation is *approximately* normal distributed.

4. Evaluate the coverage of the confidence interval for 1/lambda: mean(X) +/- 1.96S / sqrt(n).

```
n <- 40
nosim <- 1000
lambda_list <- seq(0.05, 10, by = 0.05)
coverage <- sapply(lambda_list, function(iter_lambda) {
  runs <- matrix(rexp(nosim * n, rate = iter_lambda), nosim)
  runs_mean <- apply(runs, 1, mean)
  runs_sd <- apply(runs, 1, sd)
  ll <- runs_mean - 1.96 * runs_sd / sqrt(n)
  ul <- runs_mean + 1.96 * runs_sd / sqrt(n)
  mean(ll < 1/iter_lambda & ul > 1/iter_lambda)
})
coverage[lambda_list == 0.2]
```

```
## [1] 0.924
```

We use simulation to test the coverage of confidence interval on differenct value of lambda. Specially, the coverage is 0.924 when lambda is set as 0.2

From Figure 3 we found that the coverage consistently fluctuates around 92.5%, regardless of change the lambda.

Appendix

Figure 1. Distribution of Sample Means of Exponential Distribution

```
g <- ggplot(data = data.frame(samples), aes(x = samples))
g <- g + geom_histogram(fill = "salmon", binwidth = 0.25, aes(y = ..density..),
                        colour = "black")
g <- g + geom_density(size = 2)
g <- g + geom_vline(xintercept = theory_mean, size = 2)
g <- g + labs(x = "means of 40 exponential(0.2)s", y = "density")
g <- g + stat_function(fun = dnorm, colour = "green", size = 1,
                       args = list(mean = theory_mean,sd = theory_std_err))
g
```

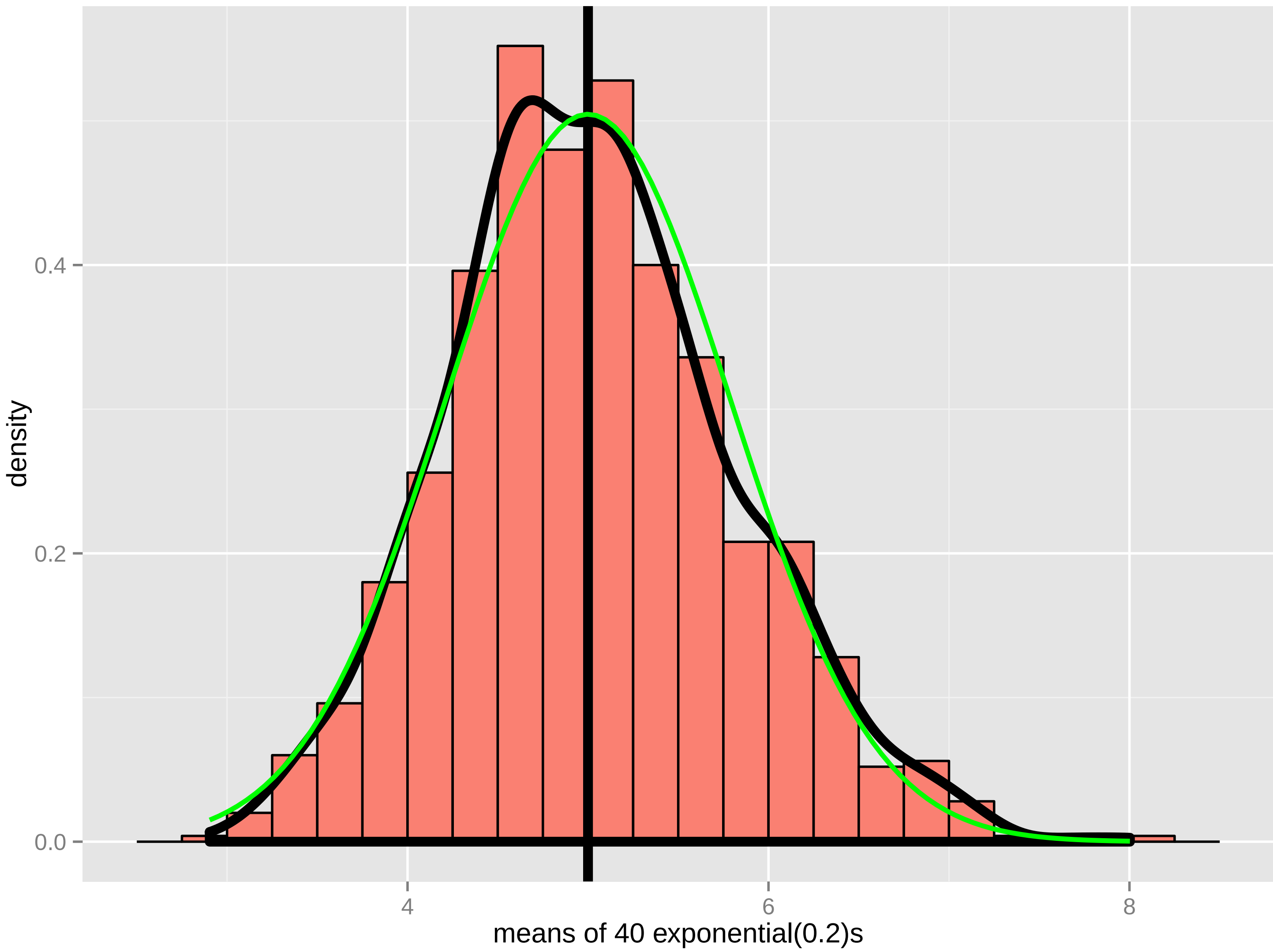


Figure 2. Q-Q Plot for sampled means of exponential distribution.

```
qqnorm(samples)
```

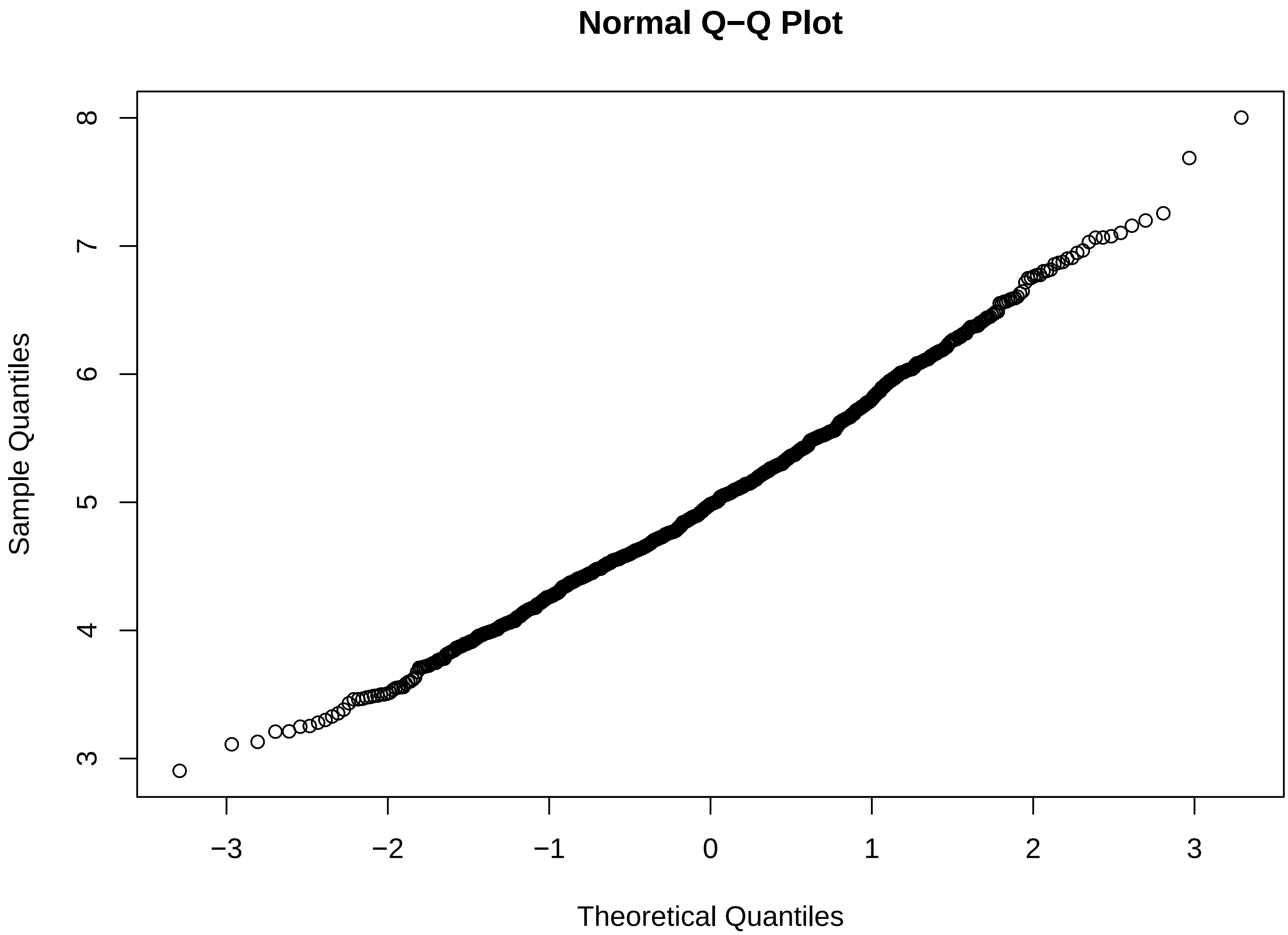
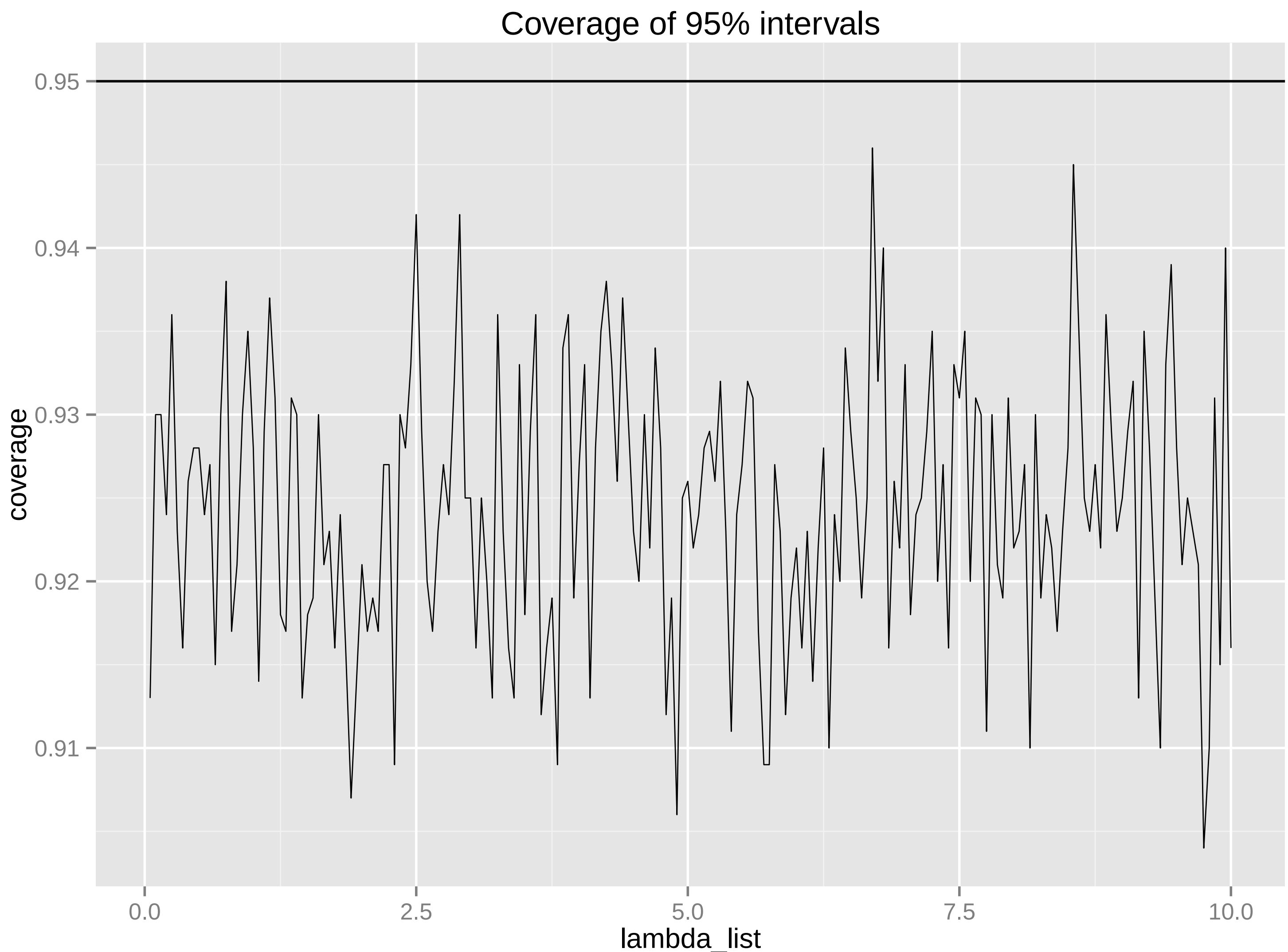


Figure 3. Coverage of the condidence interval

```
g <- ggplot(data.frame(lambda_list, coverage), aes(x=lambda_list, y=coverage))
g <- g + geom_line(size = 0.25) + geom_hline(yintercept = 0.95)
g <- g + ggtitle("Coverage of 95% intervals")
g
```


Coverage of 95% intervals
coverage
0.95
0.94
0.93
0.92
0.91
0.0
2.5
5.0
7.5
10.0
lambda_list