Statistical Inference Course Project

Intoduction

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. In this simulation, we will investigate the distribution of averages of 40 exponential(0.2)s.

```
N = 40 # number of exponential
lambda = 0.2
n_repetitions = 3000
theoretical_mean = 1/lambda
theoretical_sd = 1/lambda

simulations = replicate(n_repetitions,rexp(N,lambda))
simulated_means = apply(simulations,2,mean) # distribution of the mean of 40 exponentials
```

Distribution of simulated means

```
center_of_simulated_means = mean(simulated_means)
cat("Theoretical mean:", theoretical_mean,  "\nCenter of simulated means:", center_of_simulated_means)
```

```
## Theoretical mean: 5
## Center of simulated means: 4.991
```

```
theoretical_sd_for_sim_means = theoretical_sd/sqrt(N)
real_sd_simulated_means = sd(simulated_means)

cat("Theoretical S.D. of simulated means:", theoretical_sd_for_sim_means,  "\nReal S.D. of simulated mea
```

```
## Theoretical S.D. of simulated means: 0.7906
## Real S.D. of simulated means: 0.8085
```

We can see that theoretical mean and theoretical S.D. for simulated means are very near to the same characteristics of empirical distribution of means. This distribution of simulated means is shown on Fig. 1 in appendix. It looks like approximately normal.

```
shapiro.test(simulated_means)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  simulated_means
## W = 0.9909, p-value = 7.27e-13
```

```
shapiro.test(log(simulated_means))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  log(simulated_means)
## W = 0.9992, p-value = 0.177
```

But Shapiro test shows normality only for log-transformed distribution (p-value>0.05).

95%-coverage

Here we calculate 95% theoretical confidence interval:

```
lo_ci = simulated_means -1*qnorm(0.975)*apply(simulations,2,sd)/sqrt(N)
hi_ci = simulated_means +1*qnorm(0.975)*apply(simulations,2,sd)/sqrt(N)
coverage = mean(theoretical_mean>lo_ci & theoretical_mean<hi_ci)

cat("Coverage of 95% confidence interval:",coverage)
```

```
## Coverage of 95% confidence interval: 0.9203
```

We can conclude that theoretical 95% coverage in this case is rather good approach. Real coverage is approximately 92%. Also see Appendix, Fig. 2.

Conclusion

We can see that theoretical means almost exactly equal to center of distribution of simulated means. Theoretical S.D. of simulated means near the real S.D. of simulated means. Distribution is looks like normal. Coverage of 95% confidence interval is 92%.

Appendix

Fig 1. Distribution of simulated mean

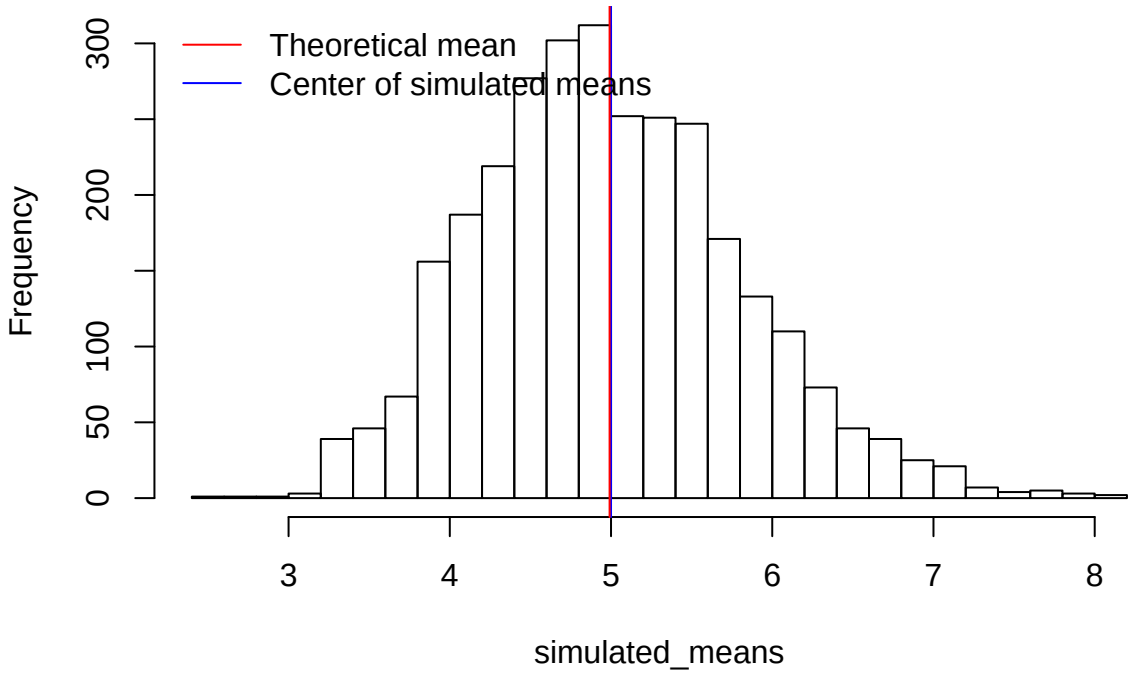

Fig 2. Coverage of 95% confidence interval