Statistical inference project. Part 1.

Anton

Wednesday, September 03, 2014

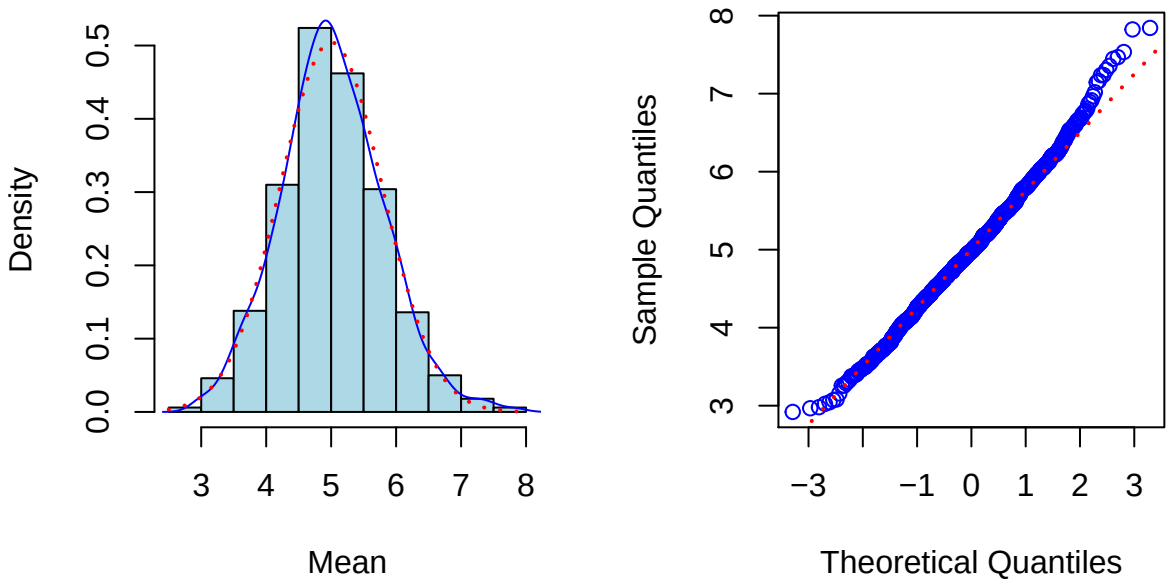
Simulation exercises

The exponential distribution simulated with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is $1/\lambda$ and the standard deviation is also $1/\lambda$. For all of the simulations λ set to 0.2. Through 1000 simulation of averages of 40 exponential we try to answer four questions:

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution. The distribution of sample means is centered at 5.016 while the theoretical center of the distribution is 5. So it's very close to the theoretical center of the distribution.

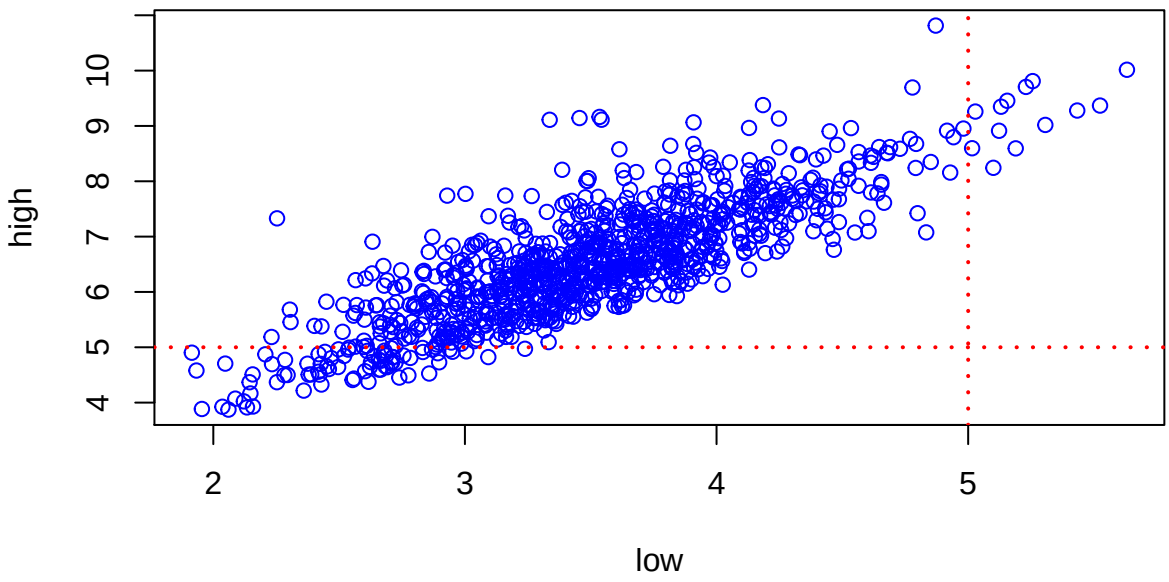
2. Show how variable the distribution and compare it to the theoretical variance of the distribution. The sample variance is 0.6262 while the theoretical variance is 0.625. Also standart deviation is 0.7913 for the sample distribution and 0.7906 for the theoretical distribution.

3. Show that the distribution is approximately normal. Compare our simulated distribution and standart normal distribution by plotting scaled simulated means.
Also compare two distributions (simulated and normal) by plotting their quantiles against each other.



So on left figure we can see that our simulated distribution pretty close to normal distribution. The Q-Q plot (right figure) shows that most of the data points are on or near the straight line, suggests that the data is almost normally distributed.

4. Evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$. The 95% confidence interval is 4.9669, 5.065 for simulated distribution and the true mean is 5.



The coverage of all simulated confidence intervals is only 91.1%.

Appendix

```
lambda = 0.2
n = 40
nsim = 1000
tmean <- 1/lambda
tsd <- (1/lambda)/sqrt(n)
tvar <- tsd^2

set.seed(5639)
sim <- matrix(rexp(nsim*n, lambda), ncol = n)
exp_means <- rowMeans(sim)
mean <- mean(exp_means)
sd <- sd(exp_means)
var <- var(exp_means)

#require(ggplot2)
par(mfrow=c(1,2))
hist(exp_means, freq = FALSE, col = "light blue", main = NULL, xlab = "Mean")
lines(density(exp_means), col = "blue")
curve(dnorm(x, tmean, tsd), col="red", lty = 3, lwd = 2, add = TRUE)
qqnorm(exp_means, main="", col = "blue")
qqline(exp_means, col="red", lty = 3, lwd = 2)
```

```
ci95 <- mean + c(-1,1)*qnorm(0.975)*sqrt(var)/sqrt(length(exp_means))

ci <- data.frame("low" = NULL,'high' = NULL)
for(i in 1:nsim){
    ci[i,"low"]<- mean(sim[i,])-1.96*sd(sim[i,])/sqrt(n)
    ci[i,"high"]<- mean(sim[i,])+1.96*sd(sim[i,])/sqrt(n)
    }

coverage <- mean(ci$low < tmean & tmean < ci$high)*100
plot(ci, col = "blue")
abline(h = tmean, col="red", lty = 3, lwd = 2)
abline(v = tmean, col="red", lty = 3, lwd = 2)
```

Full R code.