Statistical Inference Part I

Saturday, November 22, 2014

Simulate exponential distribution

Note:'dplyr' and 'matrixStats' are required prior to conduct analysis.

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

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

By simulating random samples with mean = 1/lamdea. Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s.

Set up parameters

The random number seed, the rate parameter `lambda`, the number of observations(sample.size)`, and the number of simulations(n).

```
set.seed(723321); lambda = 0.2; sample.size = 40; n = 1000
```

Now we take 40000 random draws from the exponential distribution. We will organize this into a matrix with 1000 rows and 40 columns and view this as simulating 1000 repetitions of a draw of 40 random exponentials with rate $\lambda = 0.2$.

```
dist <- matrix(rexp(sample.size*n, rate=lambda), ncol = sample.size, nrow=n)
```

Calcualate sample means, sample standard deviation, left and Right confidence interval end points.

```
dist.mean <- rowMeans(dist)
dist.sd <- rowSds(dist)
dist <- cbind(dist, dist.mean = dist.mean, dist.sd = dist.sd)
dist <- data.frame(dist)
dist <- dist %>%
    mutate(CI_l = dist.mean - 1.96 * dist.sd / sqrt(sample.size),
           CI_r = dist.mean + 1.96 * dist.sd / sqrt(sample.size))
```

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

The mean of the sample means is 5.0100662 versus the theoretical mean, where is

$$\frac{1}{\lambda} = \frac{1}{0.2} = 5.$$

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

The standard deviation of the simulated sample means is 0.7754508 vs. the theoretical standard deviation, where is

$$\frac{\sigma}{\sqrt{n}} = \frac{1/\lambda}{\sqrt{40}} = 0.7905694.$$

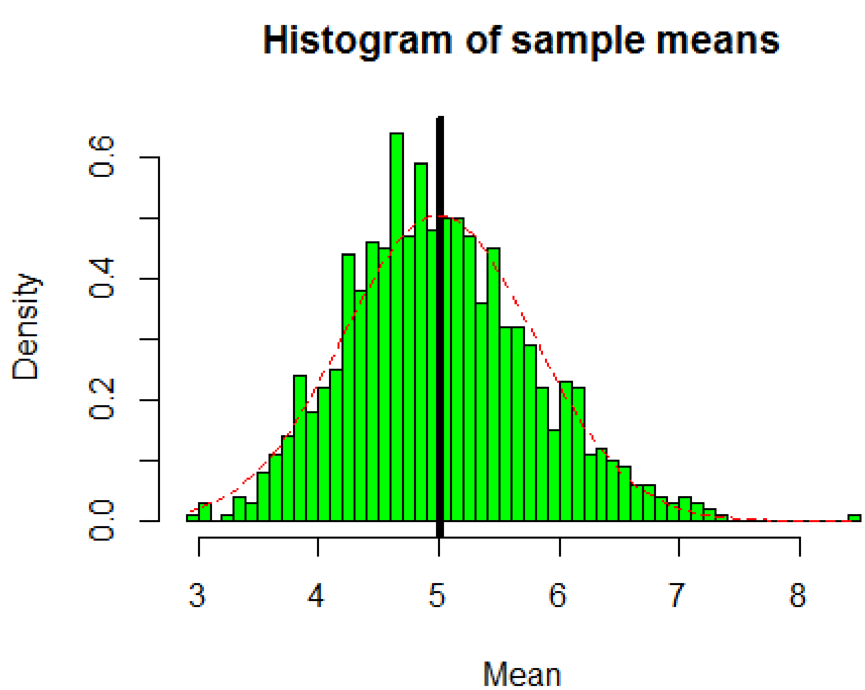
The variance of our sample means is 0.6013239 versus the theoretical variance, where is

$$\frac{\sigma^2}{n} = \frac{1/\lambda^2}{40} = 0.625.$$

3. Show that the simulated data distribution is approximately normal.

The histogram below illustrates the distribution of sample means is closely approaching the normal distribution of $N\left(1/\lambda, \frac{1/\lambda}{\sqrt{40}}\right)$.

```
h <- hist(dist$dist.mean, breaks=sample.size, prob=T, col="green",
          main = "Histogram of sample means",
          xlab = "Mean")
abline(v=mean(dist.mean),col="black",lwd=4)
xfit <- seq(min(dist.mean), max(dist.mean), length=100)
yfit <- dnorm(xfit, mean=1/lambda, sd=(1/lambda/sqrt(sample.size)))
lines(xfit, yfit, pch=22, col="red", lty=2)
```



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

$$\bar{x} \pm 1.96\frac{s}{\sqrt{n}}$$

```
dist <- dist %>%
    mutate(Conf.I = (CI_l < 1/lambda) & (1/lambda < CI_r))
coverage <- mean(dist$Conf.I)

print(coverage)

## [1] 0.937
```

The confidence inverval of simulated sample date capture 93.7% of the true mean $\frac{1}{\lambda} = 5$.