Statistical Inference - Course Project - Question 1

2014-09-21

Synopsis

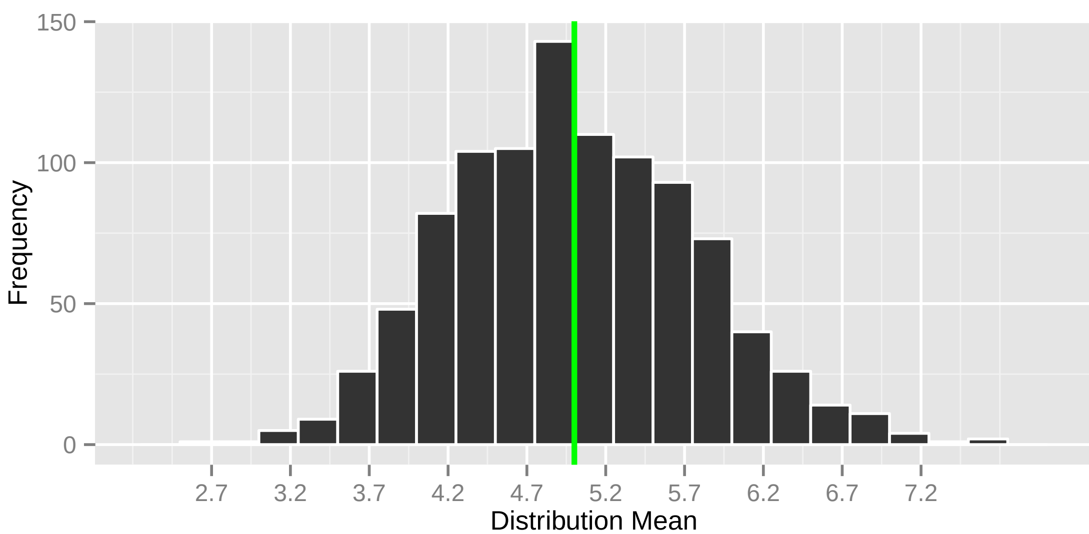
In this simulation, I will investigate the distribution of averages of 40 exponential(0.2)s. I will illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. Initilize our simulation dataset with λ=0.2, 40 samples and 1000 rows per sample.

```
lambda <- 0.2; s <- 40; n <- 1000;
dist <- matrix(rexp(s*n, rate=lambda), ncol=s, nrow=n)
distMean <- rowMeans(dist)
```

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

Calculate the expected theoretical center and the center (mean) of `distMean`.

```
centerTheoretical <- 1/lambda; centerCalculated <- mean(distMean);
```



The theoretical center of the distribution is: $1/\lambda = 1/0.2 = 5$. Our calculated center of the distribution, indicated by the green line, is 5.001. The two are very close.

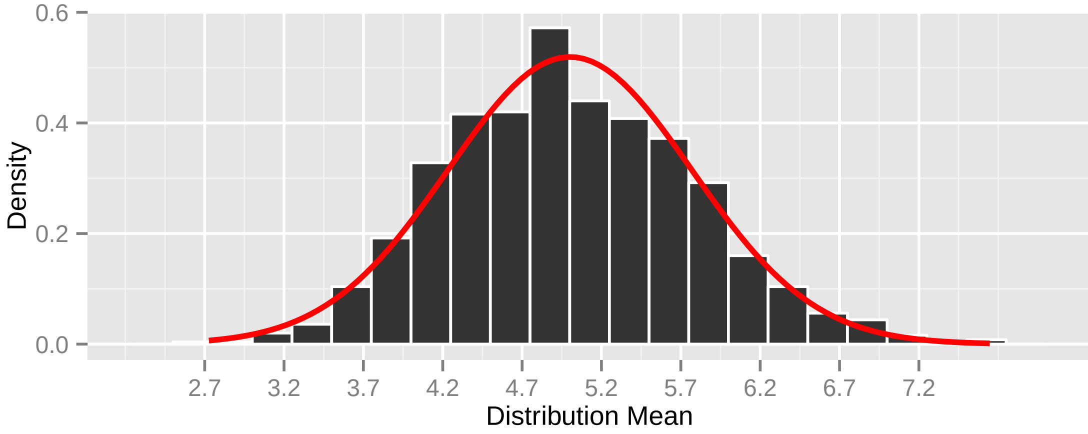
2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
sdTheoretical <- (1/lambda)*(1/sqrt(s)); sdCalculated <- sd(distMean);
varTheoretical <- sdTheoretical^2; varCalculated <- sdCalculated^2;
```

The theoretical standard deviation is calculated as follow: $(\frac{1}{\lambda} * \frac{1}{\sqrt{n}}) = 0.7906$ yielding a variance of 0.625. Our calculated standard deviation of the distribution is 0.7681 yielding a calculated variance of 0.59. Again the theoretical and calculated values are very close to each other.

3. Show that the distribution is approximately normal.

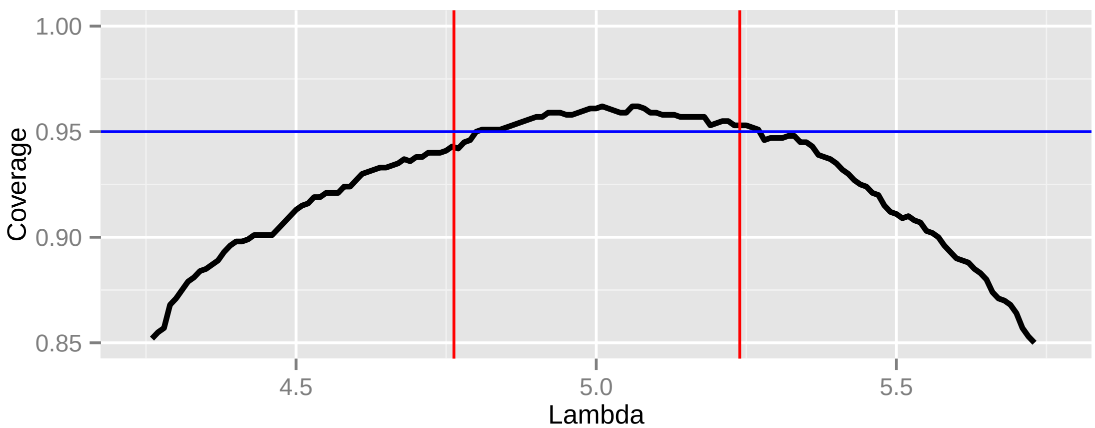
In an effort to compare our calculated distribution with a normal distribution I render the histogram shown earlier with the Y Axis showing the density as opposed to the previous shown frequency and I then overlay a Normal Distribution (red line) using the our `centerCalculated` and `sdCalculated` variables. This plot clearly shows that our distribution is approximately normal.



4. Evaluate the coverage of the confidence interval for $1/\lambda = \bar{X} \pm 1.96\frac{S}{\sqrt{n}}$.

Start by creating a function `calcCover` that returns the coverage for a specified lambda, then create a range of lambdas, calculate their respective coverage and lastly calculate the vertical bounds (`vBounds`). Finally plot the `cover` against the `lambdas`.

```
calcCover <- function(x) {
  offset <-  (1.96 * sqrt(1/lambda**2/s));
  ll <- distMean - offset; ul <- distMean + offset;
  mean(ll < x & ul > x)
}
lambdas <- seq(4.25, 5.75, by=0.01); cover <- sapply(lambdas, calcCover);
vBounds <- centerCalculated + (c(-1,1) * 1.96 * sdCalculated/sqrt(s))
```



The horizontal blue line marks the 95% confidence interval and the vertical red lines (4.763 and 5.239) marks the above 95% interval.