Statistical Inference Part 1 - Simulation Exercise

Zorina Alliata

Tuesday, February 17, 2015

Synopsis

The exponential distribution can be simulated in R with rexp(n, lambda) where n is the number of observations and lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also also 1/lambda. In these simulation exercises, we investigate the distribution of averages of 40 exponentials over a thousand observations (n=100), assuming the lambda = 0.2

Results

Create a thousand simulated averages of 40 exponentials, i.e. rexp(40,0.2)

```
expdist <- rep(NA,1000)
for (i in 1:1000){
    expdist[i] <- mean(rexp(40,0.2))
}
```

The theoretical center of the mean is 1/lambda = 1/0.2: 5

```
calcmean <- mean(expdist)
calcmean
```

```
## [1] 5.044867
```

The calculated mean is 5.044 and the theoretical mean is 5, so the variation is negligible.

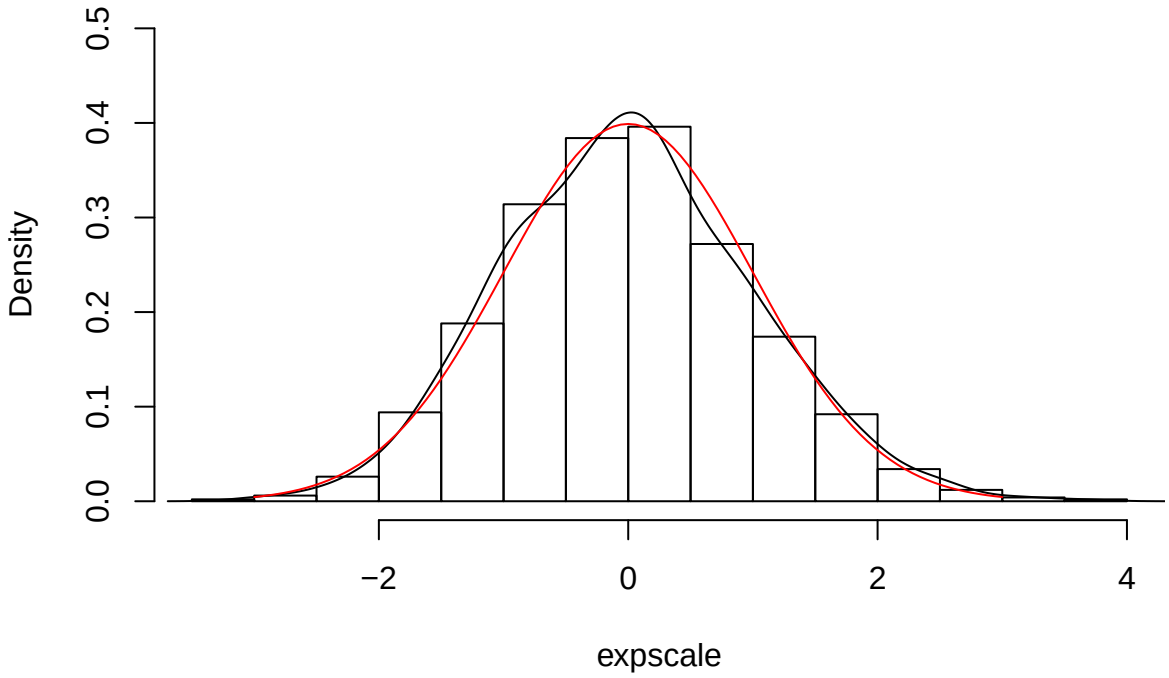
The theoretical variance is ((1/0.2)^2)/40 : 0.625

```
calcvar <- var(expdist)
calcvar
```

```
## [1] 0.6335581
```

The calculated variance is 0.6333 and the theoretical variance is 0.625, so both distributions have similar variability.

We use the scale() function to plot the distribution and compare it to a normal distribution.



This is to be expected because of the CLT.

Evaluating the coverage of the confidence interval

A 95% confidence interval should contain, if we simulate a big number of them, the mean value for the exponential distribution $(1/\lambda)$ 95% of the time.

```
set.seed(567)

lambda <- 0.2

# checks for each simulation if the mean is in the confidence interval

inconfint <- function(lambda) {

                ehats <- rexp(1000, lambda)

                se <- sd(ehats)/sqrt(1000)

                ll <- mean(ehats) - 1.96 * se

                ul <- mean(ehats) + 1.96 * se

                (ll < 1/lambda & ul > 1/lambda)

        }
```

```r
# estimate the coverage in each round of simulations
coverage <- function(lambda) {
    covvals <- replicate(100, inconfint(lambda))
    mean(covvals)
}
# perform the simulation
simres <- replicate(100, coverage(lambda))
mean(simres)
```

```
## [1] 0.9484
```

The confidence interval contains the theoretical value 94.84% of the time (close to the expected 95%).