Coursera Statistical Inference Project (Part 1): Simulation to Explore Inferences

mjurado

Thursday, March 12, 2015

Overview The goal of this analysis is to explore inferences in order to compare the sample means versus the theoretical mean of the distribution as well as compare the sample variance versus the theoretical variance of the distribution. In the end the analysis will show that the sample distribution is approximately normal.

Given Information We are given the following information:

- There should be 40 exponentials per distribution
- The default lamba value is .02 for all simulations
- The theoretical mean equals 1/lambda
- The standard deviation of the theoretical mean is also equal 1/lambda
- There should be 1000 simulations applied

Simulations

```
# Set given information
n <- 40 # number of exponentials
lambda <- .2 # rate
theo.mean <- round(1/ lambda,3) # <- theoretical mean of the distribution
theo.std.dev <- theo.mean
theo.std.err <- theo.std.dev/sqrt(n)
sim.cnt.small <- 10
sim.cnt.big <- 1000

# Calculation the sample mean for 10 simulations
sample.sim.small = NULL
for (i in 1 : 10) sample.sim.small[i] = mean(rexp(n,lambda))
sample.mean.small <- round(mean(sample.sim.small),3)

# Calculate the sample mean over 1000 simulations
sample.sim.big = NULL
for (i in 1 : 1000) sample.sim.big[i] = mean(rexp(n,lambda))
sample.mean.big <- round(mean(sample.sim.big),3)
```

Analysis 1: Mean comparisons

Show the sample mean and compare it to the theoretical mean of the distribution.

- The sample center of distribution for 10 simulations is: 4.49.
- The sample center of distribution for 1000 simulations is: 5.035.

- The theoretical center of distribution is: 5.

Conclusion: I did two sample simulations to show the effect on the mean as more simulations occurred. As you can see, as the number of simulations increased the center of distribution gets closer to the theoretical center of distribution.

```
# Determine the big and small sample variances
sample.var.small <- sample.mean.small^2
sample.var.big <- sample.mean.big^2

# Determine the variance of the theoretical distribution
theo.var <- theo.std.dev^2
```

Analysis 2: Variance comparisons

Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution. The variance of an exponential distribution is the square of the standard deviation (which happens to equal the mean) so we end up with the following:

- The small sample variance is: 20.1601.
- The large sample variance is: 25.351225.
- The theoretical variance is: 25.

Conclusion: As you can see, the greater the number of simulations, the closer the variances gets to the theoretical variance.

Analysis 3: Show that the distribution is approximately normal. Conclusion: From the plots below you can see that the sample with the 1000 simulations is much closer to a Normal distribution than its counterpart with only 10 simulations.

```
library(ggplot2)
library(gridExtra)
```

```
## Loading required package: grid
```

```
dist10 <- qplot(sample.sim.small,
              geom = "density",
              xlim= c(min(sample.sim.small)-1,max(sample.sim.small)+1),
              xlab = "",
              ylab = "",
              main="Sample With 10 Simulations vs Normal") +
            stat_function(fun = dnorm,
                      colour = "red",
```

```
                    args=list(mean=theo.mean, sd=theo.std.err))

dist1000 <- qplot(sample.sim.big,
              geom = "density",
              xlim= c(min(sample.sim.big),max(sample.sim.big)),
              xlab = "",
              ylab = "",
              main="Sample With 1000 Simulations vs Normal") +
            stat_function(fun = dnorm,
                  colour = "red",
                  args=list(mean=theo.mean, sd=theo.std.err))
grid.arrange(dist10, dist1000, ncol = 1)
```