Coursera Statistical Inference Project - Part 1

Father Abraham

13/09/2014

Part 1

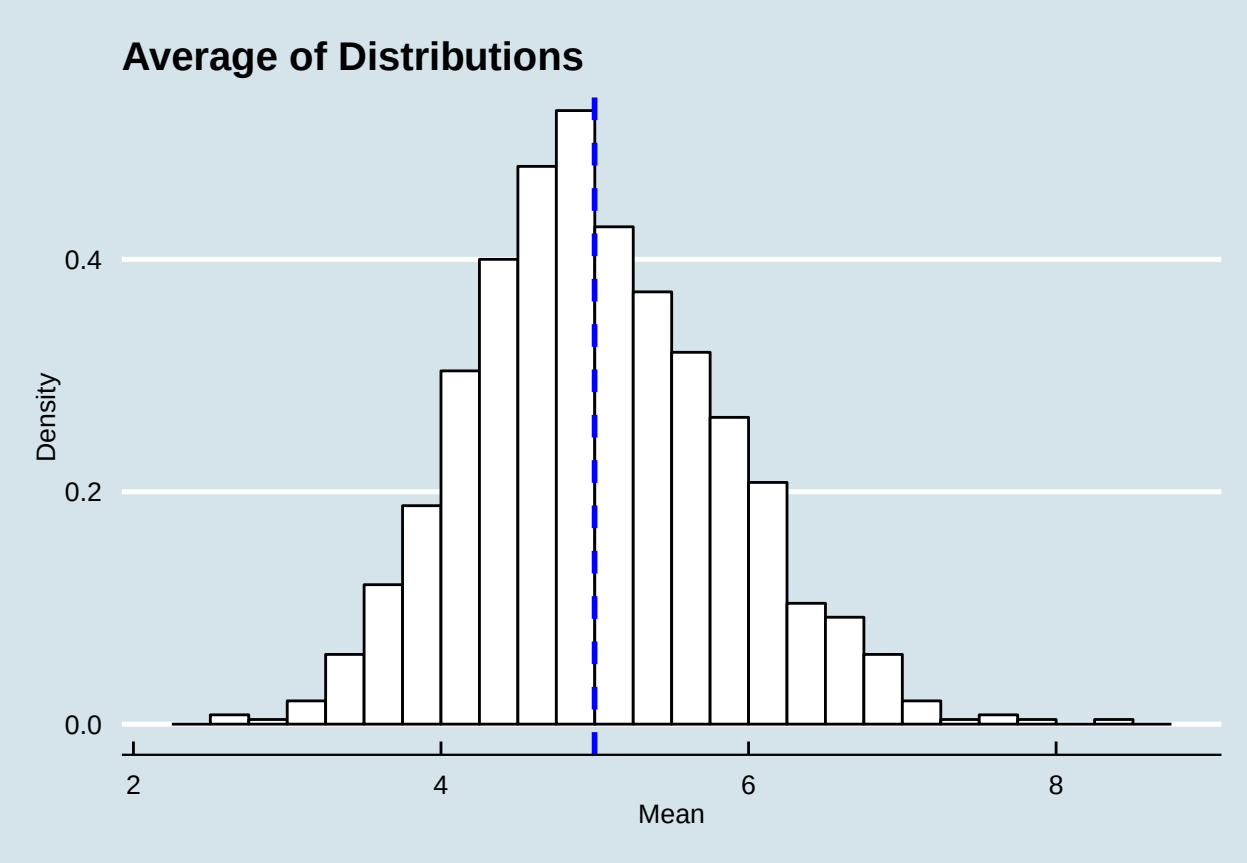
The first task of the project is to "[i]llustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s".

```
##set seed for predictable random numbers (!)
set.seed(23)

##satisfy project requirements
sims=1000
lambda=0.2
obs=40

##load and format data
dat1000 <- as.data.frame(replicate(sims, mean(rexp(obs, lambda))))
names(dat1000) <- c("Means")
```

With the data in place we will now plot the distribution as a histogram using the following code, and we add a line to show where the distribution is centered - very close to 5.:



Let's have a look at some other properties of the original distribution.

The variance:

```
var(dat1000$Means)
```

```
## [1] 0.6862
```

The theoretical variance is:

???

The standard deviation:

```
sqrt(var(dat1000$Means))
```

```
## [1] 0.8284
```

Next, we plot the normal distribution on top, we know the mean is 5 from 1/lambda, so we will use that (i.e. we won't *fully* standardize, by centering the mean on 0), and for a standard normal distribution the standard deviation is 1 (Source: http://en.wikipedia.org/wiki/Normal_distribution)

```
g <- g + stat_function(fun = dnorm, colour = "red", arg = list(mean=5, sd =1))
g
```

Clearly, the distribution is very close to a normal distribution, and we have demonstrated that the Central Limit Theorem is not refuted here.

I don't yet know how to create the 95% Confidence Level....