The Exponential Distribution

Julian Hatwell

#### Notes:

The code in this document requires the following R packages:

library(lattice)

## Contents

* Overview
* Simulations
* Sample Mean vs Theoretical Mean
* Sample Variance vs Theoretical Variance
* Appendix: Comparing Exponential and Normal Distributions

## Overview

This document will use the Exponential distribution to investigate the Central Limit Theorem, which states that the *distribution of averages* of any iid variables becomes that of a standard normal as the sample size increases, no matter what the distribution of the population.

The Exponential distribution makes a good case study as it has very different properties to the Normal distribution.

This investigation uses simulations carried out in the R programming environment.

## Simulations

The following code sets up all the parameters and constants, as well as the random seed for reproducibility:

# here are all the standard parameters. Can be tweaked for more further investigations  
set.seed(10) # reprodicibility  
lambda <- 0.2 # lambda for random samples  
sim\_size <- 1000 # simulation size  
smpl\_size <- 40 # sample size, each simulation

# create some empty vectors to fill with sim data  
mn\_smpl <- numeric(sim\_size)  
var\_smpl <- numeric(sim\_size)  
# iterate over the random simulation and collect the results by vector index  
for (i in 1:sim\_size) {  
smpl <- rexp(smpl\_size, lambda)  
mn\_smpl[i] <- mean(smpl)  
var\_smpl[i] <- var(smpl)  
}  
# put everything in a data frame to make it easy to work with  
sim\_data <- data.frame(means = mn\_smpl, variances = var\_smpl, st\_devs = sqrt(var\_smpl))

### Sample Mean vs Theoretical Mean

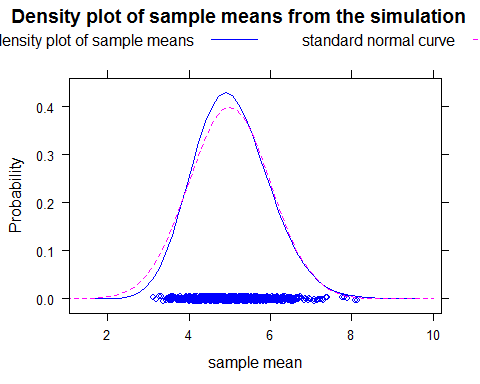
Interrogating the sample mean data shows that the mean and st. dev are indeed very close to 1/lambda (5.), which are the expected mean and st.dev for the Exponential distribution:

stats <- list(sample\_mean = mean(mn\_smpl)  
 , sample\_variance = mean(var\_smpl)  
 , sample\_st\_dev = sqrt(mean(var\_smpl)))  
stats

## $sample\_mean  
## [1] 5.04506  
##   
## $sample\_variance  
## [1] 25.56219  
##   
## $sample\_st\_dev  
## [1] 5.055907

The following code produces a density plot of the mean values from the simulation. Superimposed on this is a PDF of the Standard Normal distribution, shifted to centre on a mean of 5 (1/lambda).

densityplot(~means, sim\_data, bw = .5, lwd = 1.5  
 , panel = function(x,...) {  
 panel.densityplot(x,..., col = "blue")  
 x <- seq(0, 10, 0.1)  
 y <- dnorm(x, mean = 5, sd = 1)  
 panel.xyplot(x,y,..., type = "l", lty = 2, col = "magenta") }  
 , main = "Density plot of sample means from the simulation"  
 , xlab = "sample mean"  
 , ylab = "Probability"  
 , key = list(text = list(  
 c("density plot of sample means"  
 , "standard normal curve"))  
 , columns = 2  
 , lines = list(  
 col = c("blue", "magenta")  
 , lty = c(1,2))))



It is clear from the above graph that these are very similar curves.

### Sample Variance vs Theoretical Variance

This step will compare the variance and standard deviation of the sample means to the theoretical mean given as the formula s/sqrt(n).

The variance of the sample means should be close to 5/sqrt(40) = 5/6.3245553 = 0.7905694

stats <- data.frame(  
 theo\_var = (1/lambda)^2/smpl\_size  
 , theo\_sd = (1/lambda)/sqrt(smpl\_size)  
 , sim\_var = var(sim\_data$means)  
 , sim\_sd = sd(sim\_data$means))  
stats

## theo\_var theo\_sd sim\_var sim\_sd  
## 1 0.625 0.7905694 0.6372544 0.7982821

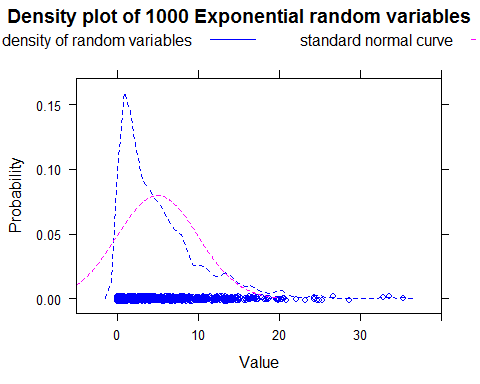
The results are very close to expectations.

### Appendix: Comparing Exponential and Normal Distributions

For completeness, more random variables are generated from the exponential function to see the difference between a large collection of random exponentials and the distribution of a large collection of averages of random variables.

The following code produces a density plot of 1000 random variables with an exponential distribution. Superimposed on this is a PDF of the Normal distribution, shifted to centre on the mean of 5 (1/lambda) and a st.dev of 5.

densityplot(rexp(1000, lambda), bw = 0.5, lwd = 1.5, lty = 2  
 , panel = function(x,...) {  
 panel.densityplot(x,..., col = "blue")  
 x <- seq(-5, 20, 0.1)  
 y <- dnorm(x, mean = 1/lambda, sd = 1/lambda)  
 panel.xyplot(x,y,..., type = "l", col = "magenta")   
 }  
 , xlim = c(-5,40)  
 , main = "Density plot of 1000 Exponential random variables"  
 , xlab = "Value"  
 , ylab = "Probability"  
 , key = list(text = list(c("density of random variables", "standard normal curve")), columns = 2, lines = list(col = c("blue", "magenta"), lty = c(1,2))))



The differences are visually apparent. From this graph, it can be seen that the Exponential distribution has the characteristic of having p = 0 for values of x < 0. The first positive value of x has p = lambda (0.2), then p drops off sharply until a long tail of higher values.

In contrast, the normal distribution is symmetrical about its mean and the standard normal has mean zero, implying negative values are just as probable as positive values.

This demonstrates the Central Limit Theorem by showing that a large sample of avarages will tend to be themselves normally distributed, even though the random variables themselves are very differently distributed.