Exponential Distribution Simulation

Michael Lee

August 21, 2015

1 Overview

The Central Limit Theorem states that the distribution of averages of independently and identically distributed variables becomes a standard normal distribution as the sample size gets larger. Here we will use randomized samples of the exponential distribution to demonstrate the CLT by comparing the theoretical mean to the sample mean and the theoretical variance to the sample variance.

2 Simulation

The exponential distribution is a probability distribution function, parametrised by λ, whose theoretical mean and standard deviation $= \frac{1}{\lambda}$. We want to show that as the sample size of the exponential distibution gets larger, the distribution of the averages of the exponential distribution approaches a normal distribution. To demonstrate this limiting behavior, we will generate a thousand averages from random exponential distributions with constant rate $\lambda = 0.2$ and sample size $n = 40$. We hold λ and n constant to satisfy the independently and identically distributed condition of the CLT.

```
n <- 1000
lambda <- 0.2
MNS <- NULL # average distribution
for (i in 1 : 1000) MNS = c(MNS, mean(rexp(40,lambda)))
hist(MNS)
```

3 Mean

The mean is a central tendency of a distribution, it represents the center of mass of a collection of locations and weights. There are two kinds of means: theoretical (μ) and sample ($\bar{x}$). The theoretical mean is calculated by integrating the expected value function and probability mass function. The sample mean is calculated by summing the arithmetic averages of empirical events.

3.1 Theoretical Mean

We know that the theoretical mean of an exponential distribution is $\mu = \frac{1}{\lambda} = \frac{1}{.2} = 5$.

3.2 Sample Mean

The Law of Large Numbers suggests that as the sample size increases, the sample mean approaches the theoretical mean. We would then expect that the sample mean of 1000 means of 40 random exponential distributions would be very close to the theoretical mean of $\mu = 5$. The following R code computes the sample mean:

```
sampleMean <- mean(MNS)
sampleMean
```

```
5.014358
```

As expected, $\bar{x} \approx \mu$.
To further illustrate this behavior, we plot the distribution of averages as a density plot and draw a vertical line over the sample mean.

```
ggplot() + aes(MNS) +
geom_density() +
geom_vline(xintercept = sampleMean, size = 1, color = 'red')
```

4 Variance

The variance of a random variable is a measure of its spread. Variance indicates the relative span and density of a set of numbers, i.e. a small variance implies a dense distribution around the mean over a short span and a large variance implies a sparse distribution around the mean over a large span. Like the mean, there are two variances: theoretical (σ^2) and sample (S^2).The theoretical variance is the expected value of the squared distance the population and its theoretical mean. The sample variance is the summation of the unbiased average squared distance between the population and the sample mean.

$\sigma^2 = E[(X-\mu)^2] = \Sigma_{i=1}^{n} \frac{(X_i-\mu)^2}{n} \mid S^2 = \Sigma_{i=1}^{n} \frac{(x_i-\bar{x})^2}{n-1}$

4.1 Sample

The following R code computes the variance of MNS and standard derivation:

```
sampleVar <- var(MNS)
std <- sd(MNS)
```

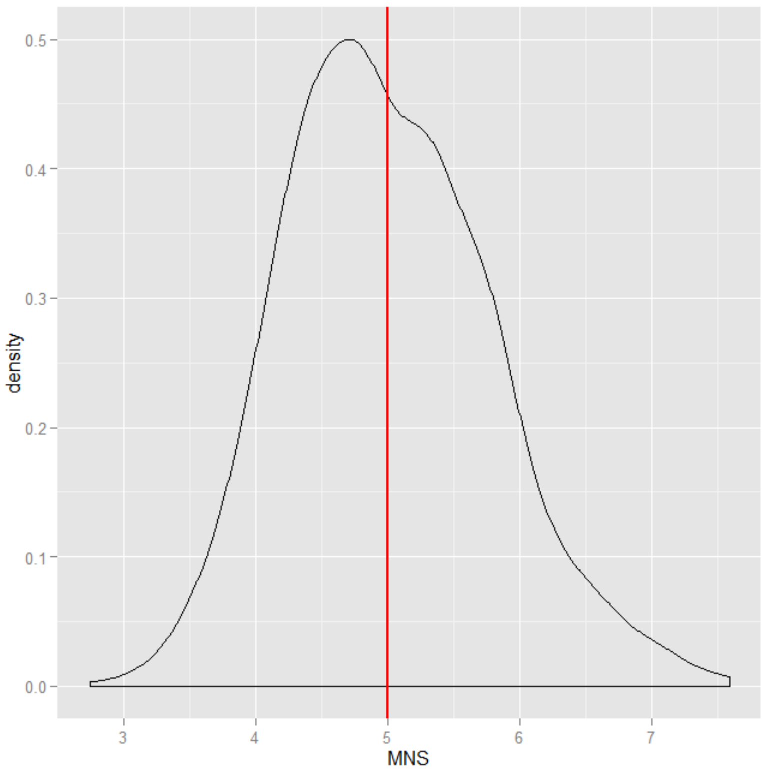

Figure 1: Density Plot of 1000 averages of $n = 40, \lambda = 0.2$ random exponential distributions. The x-intercept of the redline is the sample mean which is very close to the mean of the original exponential disitribution $\mu = 5$. Note the concentration of density surrounding the sample mean.

```
0.6021728
0.7759979
```

4.2 Theoretical

Since the sample variance is a function of the sample mean, it has an associated population distribution.

$$Var(\bar{X}) = Var(\frac{\Sigma_{i=1}^{n} X_i}{n})$$
$$Var(\bar{X}) = Var(\frac{X_1}{n} + \frac{X_2}{n} \ldots + \frac{X_n}{n})$$
$$Var(\bar{X}) = \frac{1}{n^2}(Var(X_1) + Var(X_2) + \ldots + Var(X_n))$$
$$\text{Since the sample mean is identically distributed,} Var(X_{1,2,\ldots,n}) = \sigma^2$$
$$Var(\bar{X}) = \frac{1}{n^2}(\sigma^2 + \sigma^2 + \ldots + \sigma^2)$$
$$Var(\bar{X}) = \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$
$$\text{Recall that } \sigma = \frac{1}{\lambda} = 5$$
$$Var(\bar{X}) = \frac{5^2}{40} = 0.625$$
$$std(\bar{X}) = sqrt(Var(\bar{X})) = 0.791$$

To further illustrate the spread, we plot the distribution of averages as a density plot and draw vertical lines over the sample mean and one standard deviation away from the sample mean.

```
ggplot() + aes(MNS) +
geom_density() +
geom_vline(xintercept = sampleMean, size = 1, color = 'red') +
geom_vline(xintercept = (sampleMean + std), size = 1, color = 'blue') +
geom_vline(xintercept = (sampleMean - std), size = 1, color = 'blue')
```

5 Normal Distribution

Theorem 5.1 Central Limit Theorem. *The distribution of averages of independently and identically distributed variables approaches a standard normal distribution as the sample size increases.*

The constants $\lambda = 0.2$ and $n = 40$ satisfy the condition of independently and identically distributed variables, but now we need to verify whether the distribution of averages $\cong$ standard normal.

To illustrate the normalcy of MNS, we plot MNS as a density plot as well as the non-standard normal distribution $Z \cong (\mu = \frac{1}{\lambda}, sd = \frac{sqrt(var(X))}{sqrt(n)}))$

```
mu    <- 1/lambda
sigma <- 1/lambda
```

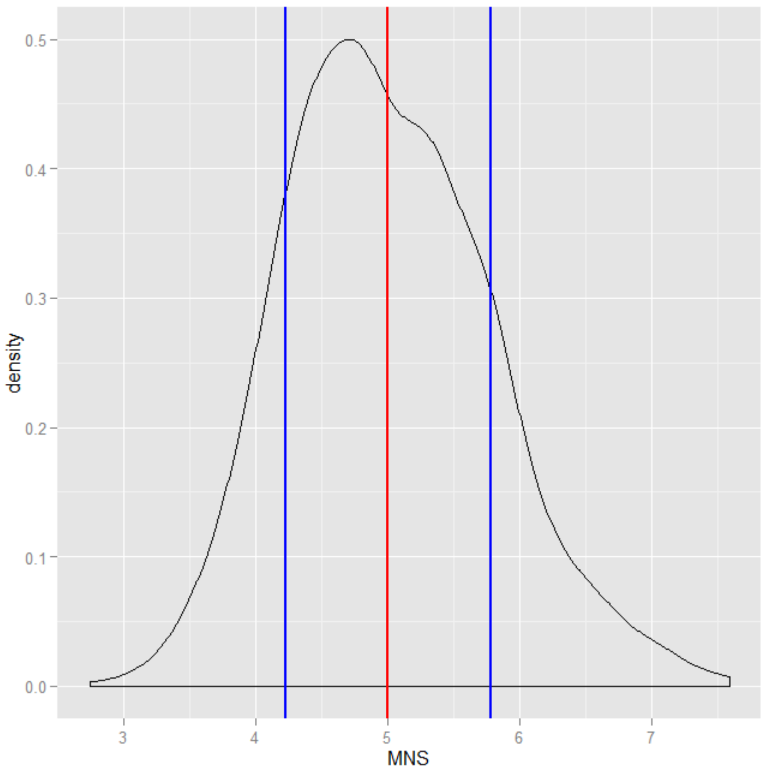

Figure 2: Density Plot of 1000 averages of $n = 40, \lambda = 0.2$ random exponential distributions. The x-intercept of the redline is the sample mean. The x-intercepts of the blue lines are one standard deviation away from the mean. Compare the distance (i.e. spread) of one standard deviation to the domain of the plot.

```
ggplot() + aes(MNS) +
geom_histogram(alpha = .10, binwidth=0.1, colour = "black", aes(y =
    ..density..)) +
stat_function(geom = "line", fun = dnorm, arg = list(mean = mu, sd =
    sigma/sqrt(40)), size = 1, colour = "green", fill = NA)
```