An Illustration of the Central Limit Theorem

Andrew Reynolds

September 22, 2015

Overview

If X is an exponential random variable with mean μ_X and variance σ_X^2, and $\bar{X}$ is the mean of an iid sample of size n, then the Central Limit Theorem says that as n becomes large, $\bar{X}$ becomes approximately normal, with mean and variance equal to

$$\mu_{\bar{X}} = \mu_X \quad \text{and} \quad \sigma_{\bar{X}}^2 = \frac{\sigma_X^2}{n}$$

respectively.

In this project, we aim to verify the above claims by simulation, using the `rexp()` function. We let $n = 40$ and take a large sample (1000 independent draws) from $\bar{X}$. If the Central Limit Theorem is true, then our sample ought to be approximately normal, with mean and variance approximately equal to the figures above (with $n = 40$). We will check this graphically and numerically.

Simulations

We let X be an exponential random variable with $\lambda = 0.2$, so that $\mu_X = \frac{1}{\lambda} = 5$ and $\sigma_X = \frac{1}{\lambda} = 5$. The following code simulates a single sample of size $n = 40$ and computes its mean. Note that `rexp()` encodes λ as `rate`.

```
set.seed(100)
example <- rexp(40, rate = 0.2)
mean(example)
```

```
## [1] 4.137412
```

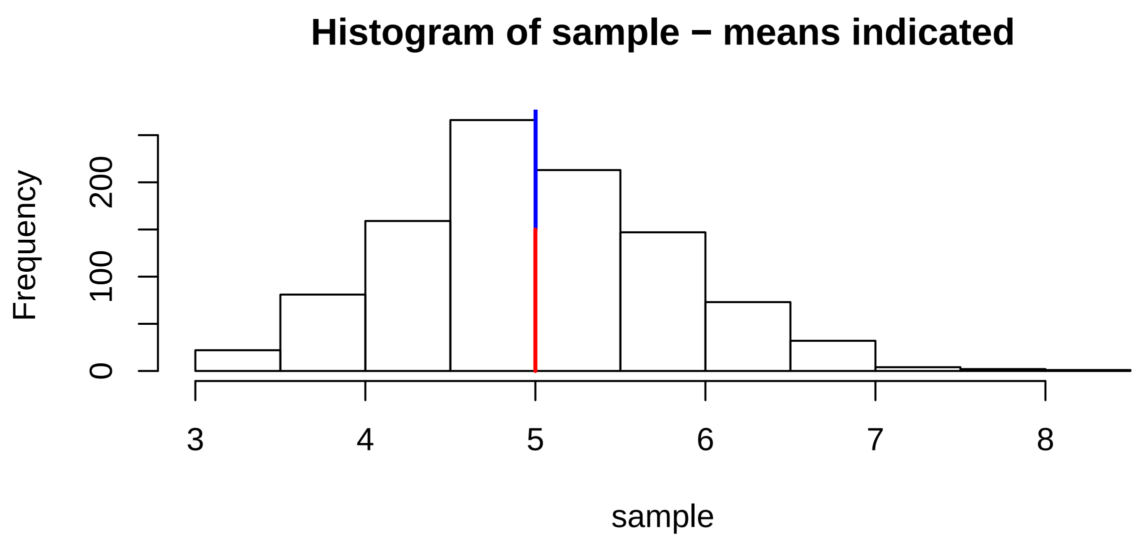
This mean is a single draw from $\bar{X}$. To obtain our sample which should represent $\bar{X}$, we repeat the above code 1000 times.

```
sample <- numeric(0)
for (i in 1:1000) {
    sample <- c(sample, mean(rexp(40, rate = 0.2)))
}
```

Sample Mean versus Theoretical Mean:

We plot the histogram of our sample and mark the sample mean with a vertical blue line, and the theoretical value with a red line.

```
hist(sample, main = "Histogram of sample - means indicated")
segments(x0 = mean(sample), y0 = 150, y1 = 300, lwd = 2, col = "blue")
segments(x0 = 5, y0 = 150, y1 = 0, lwd = 2, col = "red")
```



This figure shows that our sample mean is very close to the theoretical mean of 5. Let's check this numerically:

```
mean(sample) - 5
```

```
## [1] 0.00121979
```

The difference is small.

Sample Variance versus Theoretical Variance:

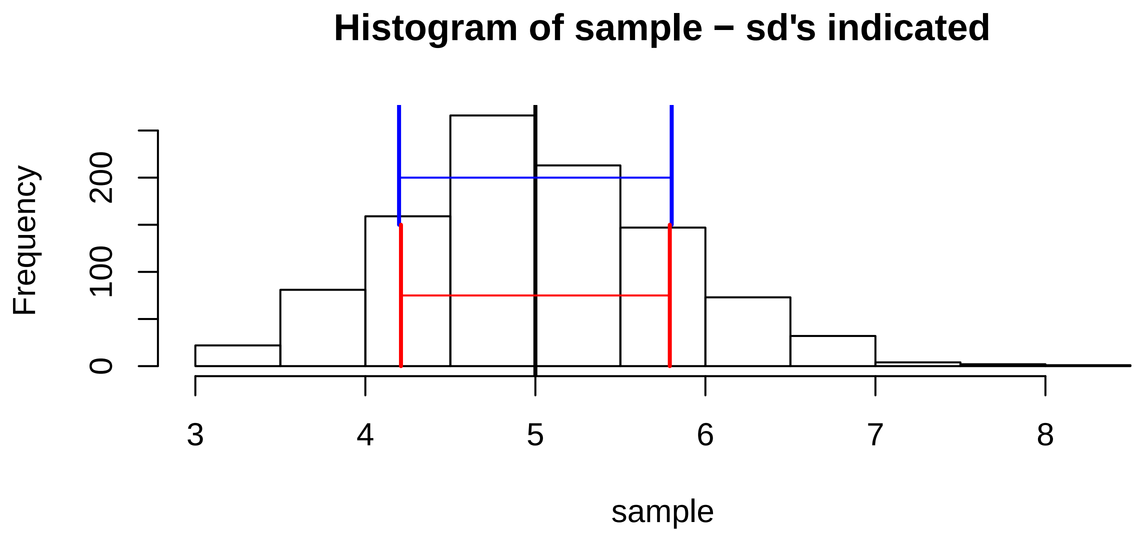
Lastly, we compare the variance of our sample with theoretical value of $\frac{25}{40}$.

```
sd(sample)^2 - 25/40
```

```
## [1] 0.01792784
```

We see that the approximation is a good one. We now include another histogram of our sample with the mean indicated (we'll use 5, since there is no visible difference between 5 and the sample mean) as well as one standard deviation from the mean in both directions. Again, blue represents the measured standard deviation and red represents the theoretical value.

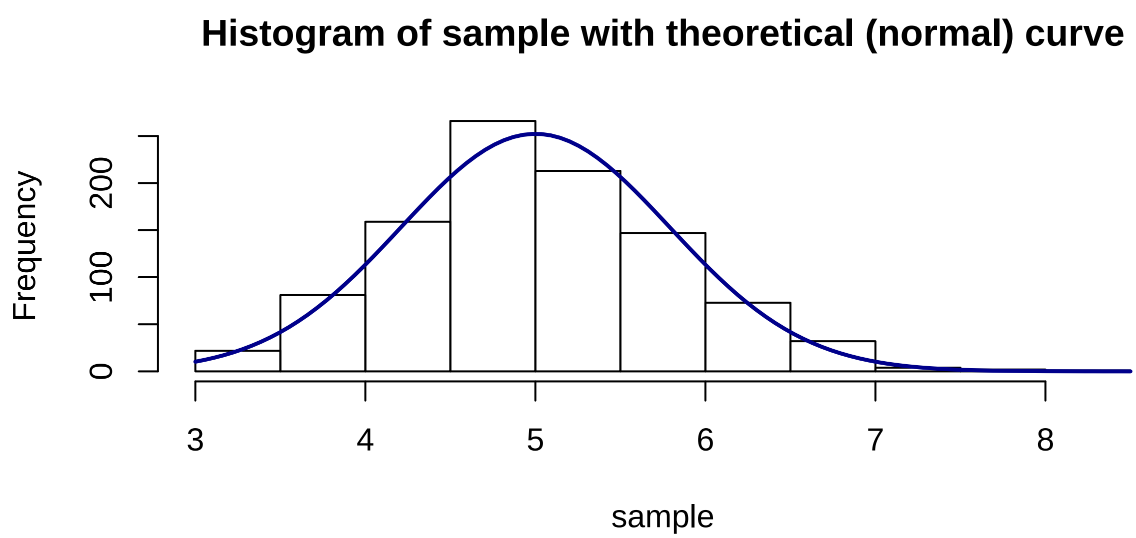
```
hist(sample, main = "Histogram of sample - sd's indicated")
abline(v = 5, lwd = 2)
segments(x0 = 5 + c(-1,1)*sd(sample), y0 = 150, y1 = 300, lwd = 2, col = "blue")
segments(x0 = 5 - sd(sample), x1 = 5 + sd(sample), y0 = 200, col = "blue")
segments(x0 = 5 + c(-1,1)*5/sqrt(40), y0 = 150, y1 = 0, lwd = 2, col = "red")
segments(x0 = 5 - 5/sqrt(40), x1 = 5 + 5/sqrt(40), y0 = 75, col = "red")
```



Distribution

Now we overlay the histogram of our sample with the a scaled version of the normal bell curve with mean and standard deviation specified by the Central Limit Theorem. Note that the height of the `dnorm` curve is multiplied by 500 to scale appropriately (the total height of the rectangles is 1000 and they have width 0.5, and so the total area is 500).

```
hist(sample, main = "Histogram of sample with theoretical (normal) curve")
curve(500*dnorm(x, mean = 5, sd = 5/sqrt(40)),
      col = "darkblue", lwd = 2, add = TRUE)
```



The histogram matches the theoretical curve very well, with a peak near the mean and roughly symmetric tails on either side.