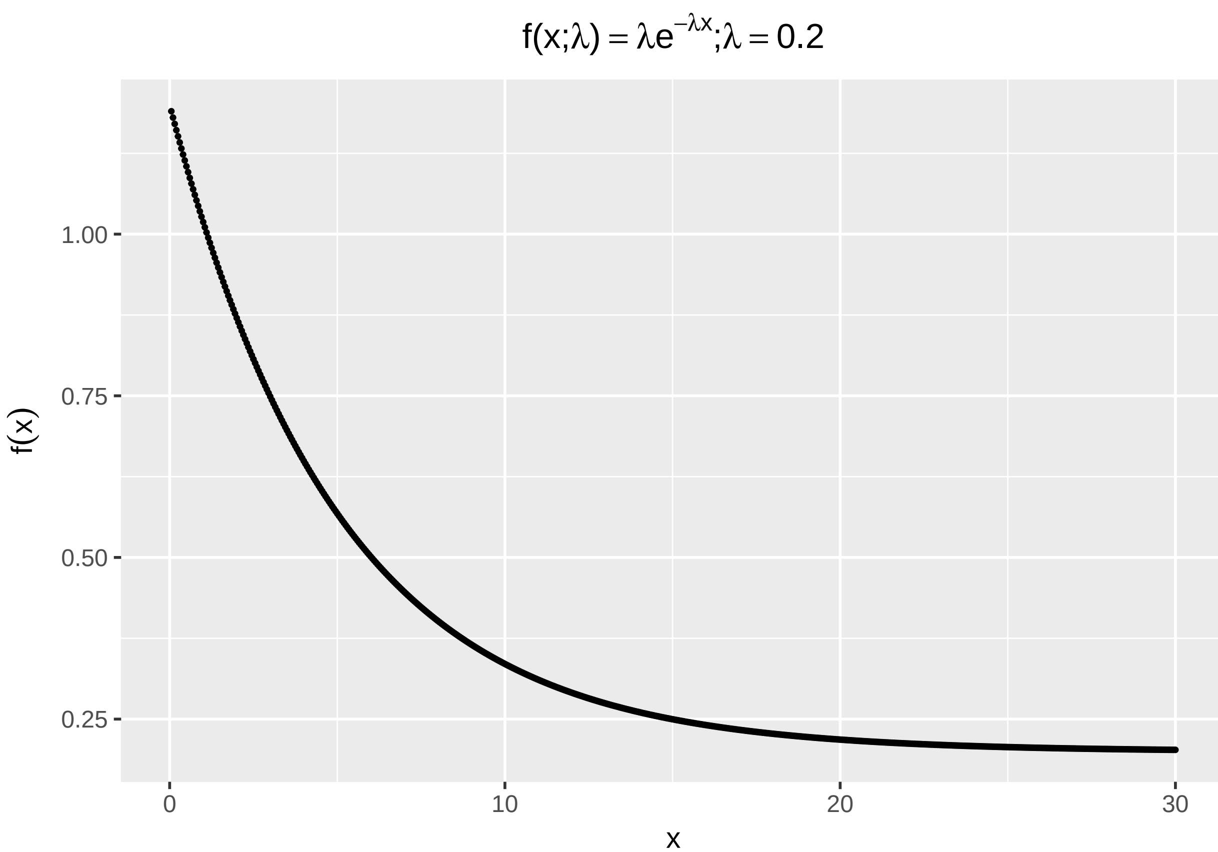
Exponential Distribution Analysis

Mike Wheelock

March 2, 2017



The Exponential Distribution

The exponential distribution is a probability distribution that describes the time between events in a Poisson process. It's formula, mean and standard deviation are:

$f(x; \lambda) = \lambda e^{-\lambda x}$

$\mu = \frac{1}{\lambda}$

$\sigma = \frac{1}{\lambda}$

If a Poisson process is telling you how many people will show up at the bus stop in a particular window at a particular average rate, the exponential distribution will give you those random wait times. This project is about investigating this distribution and illustrating the Central Limit Theorem.

Procedure

For the purposes of this exercise an exponential distribution with $\lambda = 0.2$ was chosen. A set of 1000 groups of n = 40 values each was sampled, collected and analyzed. This was accomplished by creating a 1000 x 40 matrix of random variables from the exponential distribution (using the `rexp` function) and then collecting statistical measures over each row. Below is the code that accomplises this.

```
# Set parameters
lambda <- 0.2
n <- 40
samples <- 1000

#Generate random variables
exp_rands <- rexp(n * samples, rate = lambda)
dim(exp_rands) <- c(samples, n)

# Calculate statistical measures
means <- apply(exp_rands, 1, mean)
variances <- apply(exp_rands, 1, function(vec) {
    mu = mean(vec);
    sum((mu - vec)^2) / (length(vec) - 1) })
sample_mean <- mean(means)
sample_variance <- mean(variances)
```

Results

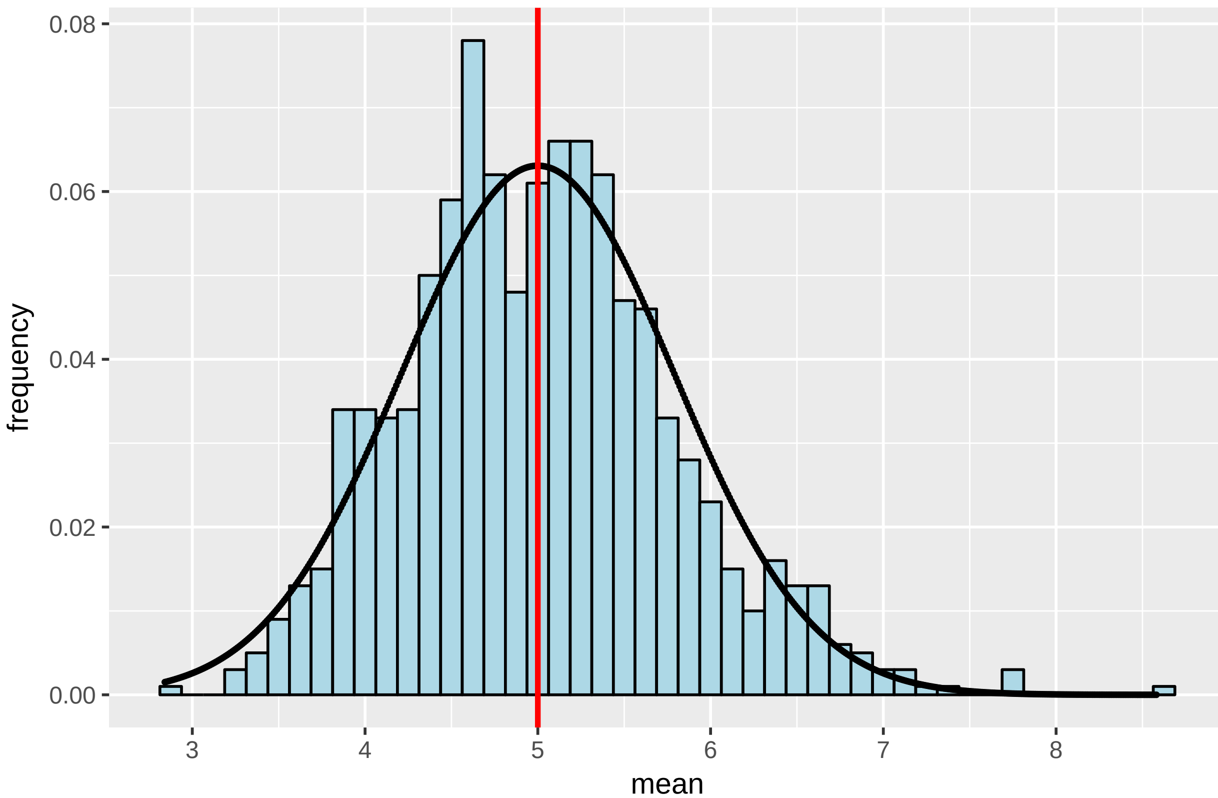
Below is a table summarizing the overall statistical measures calculated for the distribution:

	Sample	Theoretical	Percent Error
Mean	5.0100102	5	0.2002034%
Variance	25.0078069	25	0.0312277%

Both the sample mean and sample variance overall were excellent estimators of the population mean and variance. It is clear from the figure below that the sample mean distribution is normally distributed. To help illustrate this I have overlaid a standard normal curve with a mean of 5 and a standards deviation of 5 on the graph. The red line is centered at x = 5, the population mean.

The second graph shows the distribution of the sample variance. As expected the variance is skewed to the right. The red line here is centered at the population variance of 25.

Simulated sample means versus normal distribution



Simulated sample variance