The Central Limit Theorem and the Exponential Distribution

T. Bennett

Monday, May 18, 2015

Synopsis

This report shows that the central limit theorem applies for the exponential distribution, i.e. the means of sets of random exponentials tend to be normally distributed with their mean at the population mean.

Simulations

Here we generate 1000 random exponentials with $\lambda = 0.2$ and display the `x` values using a histogram with an overlain density curve.

1000 random exponentials
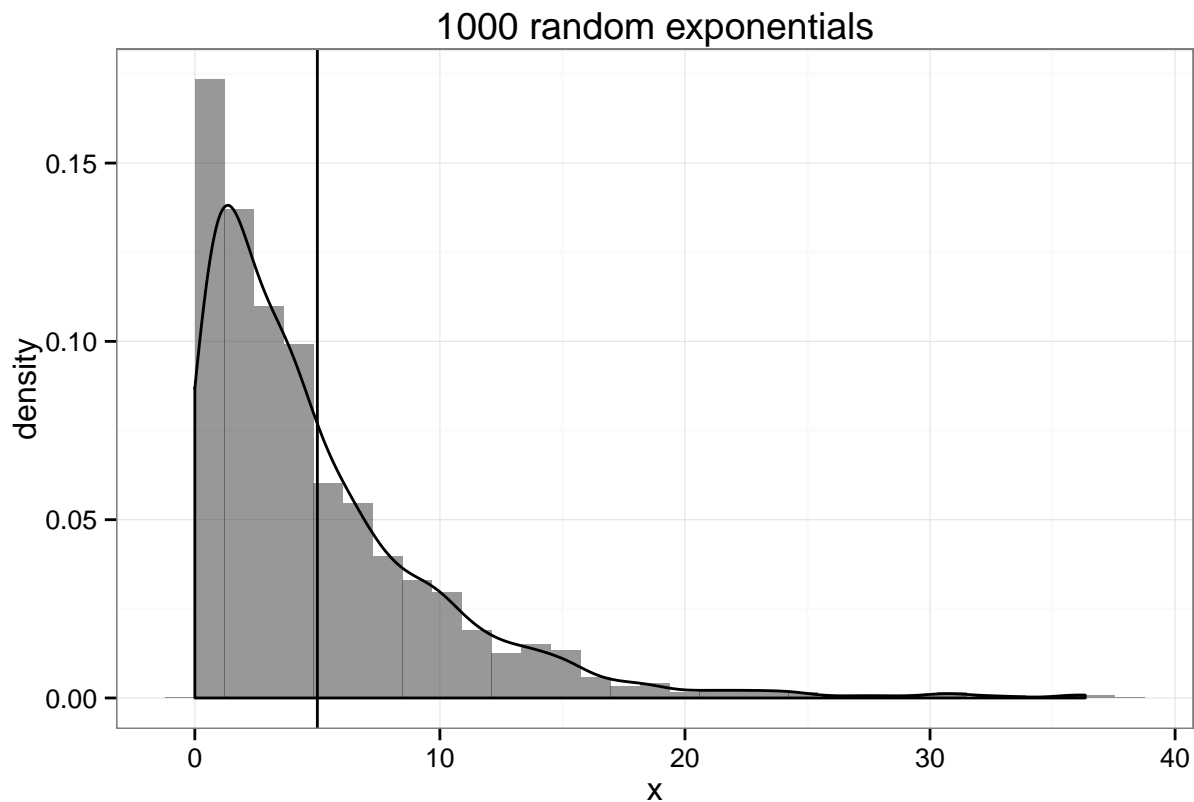
The theoretical mean of the distribution should be $1/\lambda = 5$ and the theoretical variance should be $(1/\lambda)^2 = 25$. The observed values are quite close to the theoretical values.
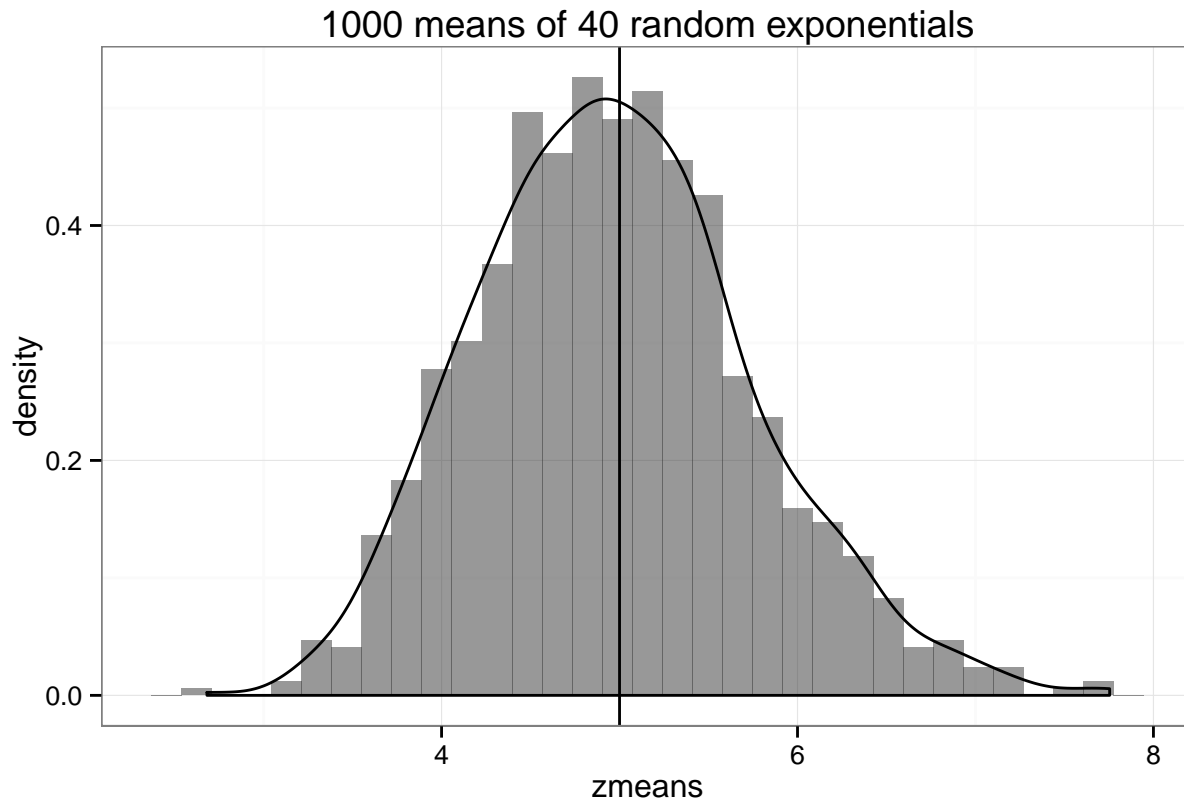
```
mean(x)
```

```
## [1] 4.991419
```

```
var(x)
```

[1] 25.53088

Here we generate 40 exponentials with $\lambda = 0.2$ **1000 times** and display the mean of each group of 40 exponentials using a histogram with an overlain density curve.

1000 means of 40 random exponentials

The theoretical mean of the distribution of 1000 means should be the population mean $1/\lambda = 5$ and the observed mean is quite close to that. The standard error of the mean should be $\sqrt{((1/\lambda)^2)/n} = 0.016$ and the sample value is just a bit larger.

```
mean(zmeans)
```

[1] 4.972126

```
sd(zmeans)
```

[1] 0.7737079

The above figure shows that the distribution of the means of 1000 groups of 40 exponentials is approximately normal, much more so than the single group of 1000 exponentials. The mean is very near the population mean, and the variance is symmetric.

Appendix

R code - run above, not run here

Set random number seed and load packages

```
set.seed(1234)
library(ggplot2)
library(dplyr)
```

Generate one set of 1000 random exponentials and show that histogram

```
x <- (rexp(1000, 0.2))
y <- as.data.frame(x)
p <- ggplot(y, aes(x = x)) + theme_bw()
p + geom_histogram(aes(y = ..density..), alpha=0.5) + geom_density(size = 0.5, alpha = 0.2) + geom_vlin
```

Generate 40 exponentials **1000 times** and show that histogram

```
nosim <- 1000
n <- 40
z <- matrix(rexp(nosim * n, 0.2), nrow = nosim)
zmeans <- apply(z, 1, mean)
zmdf <- as.data.frame(zmeans)
p2 <- ggplot(zmdf, aes(x = zmeans)) + theme_bw()
p2 + geom_histogram(aes(y = ..density..), alpha=0.5) + geom_density(size = 0.5, alpha = 0.2) + geom_vli
```