Exponential Distributions

Alex Petkovski

Saturday, May 23, 2015

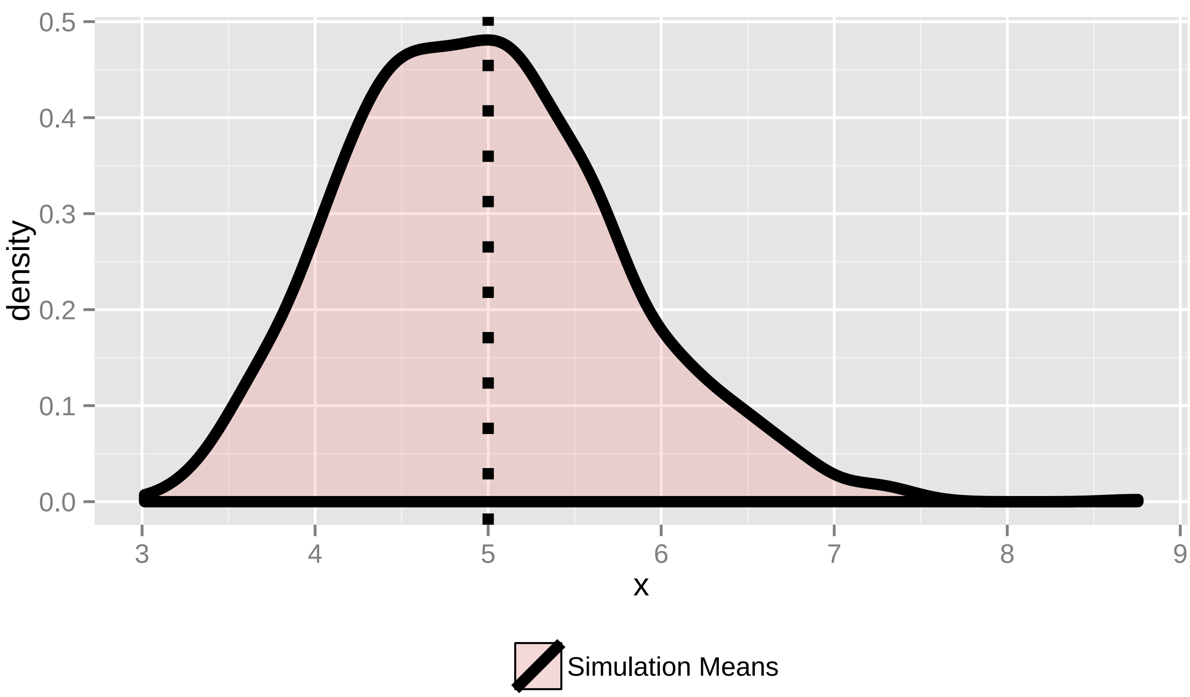
Overview

The following report investigates properties of the **Exponential Distribution** in **R** wherein we will compare it with the ***Central Limit Theorem*** and we will set **lambda = 0.2** for all simulations. Using means of 40 exponentials from a sample of 1000, we will compare **sample vs. theoretical mean** and **sample vs. theoretical variance**. Finally, we will compare the difference between a large collection of **random exponentials** and the distribution of a large collection of **averages of 40 exponentials** and show that it is approximately normal.

Simulations

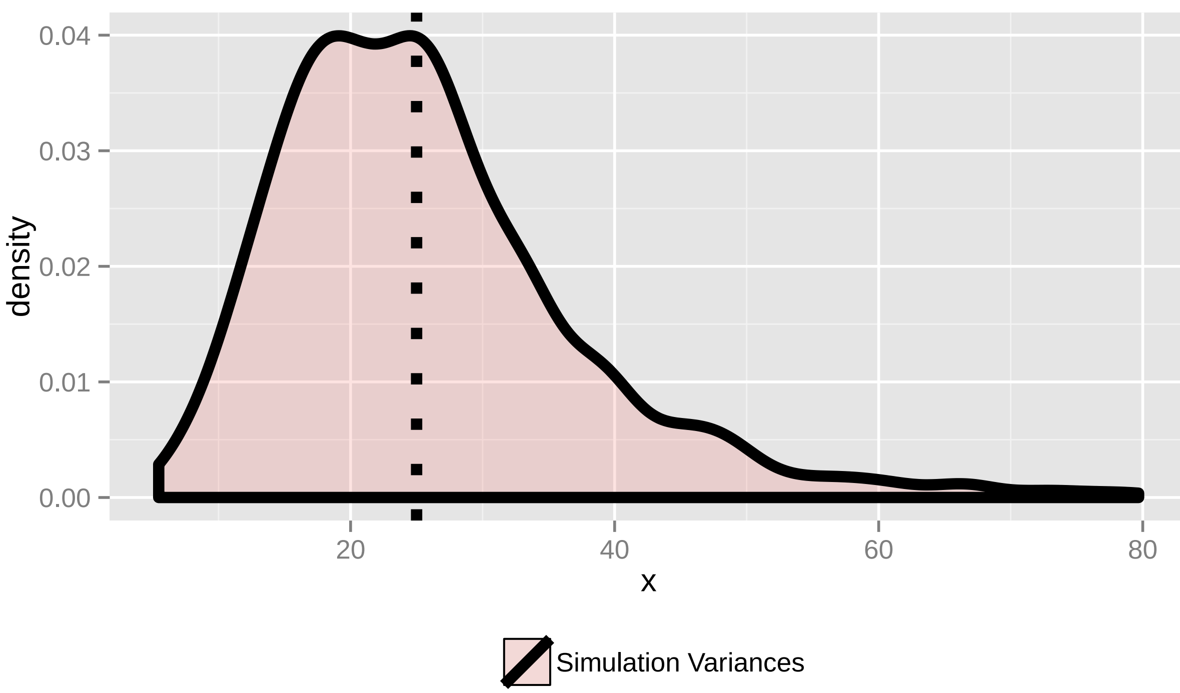
Sample Means

- Simulate $n = 40$ exponentials a 1000 times
- Let X_i be the outcome for sample i
- Then note that $\mu = E[X_i] = 1/\lambda = 1/0.2 = 5$ (Vertical dotted line)
- Let's take n exponentials, take their mean and repeat this 1000 times (Simulation Means)



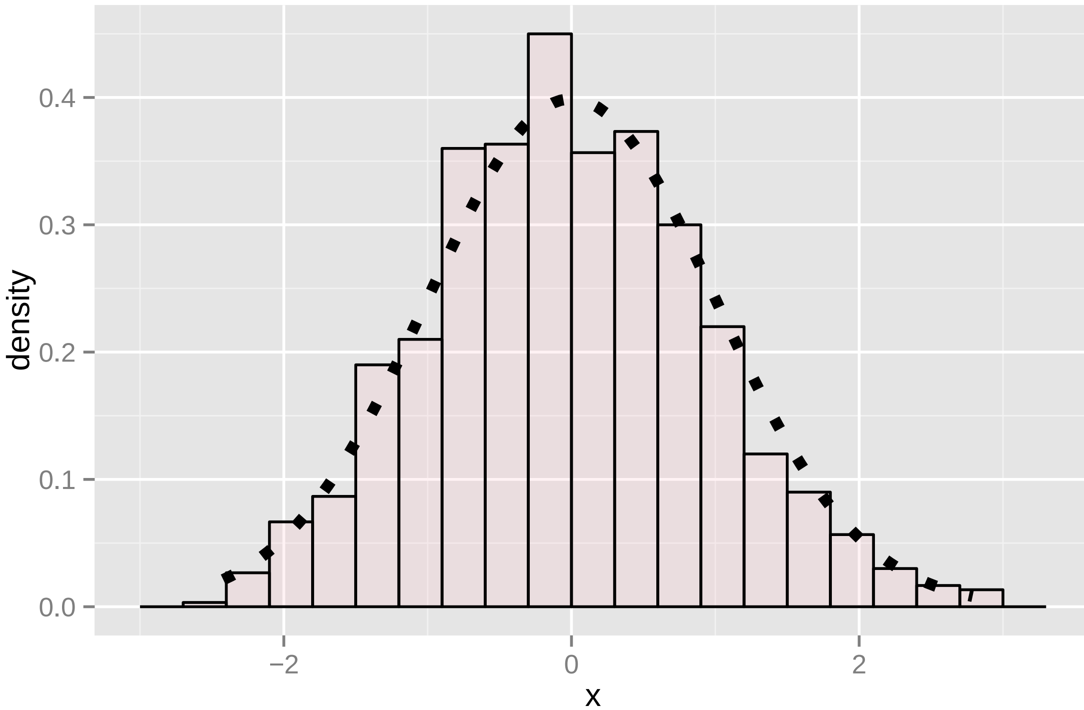
Sample Variances

- Simulate $n = 40$ exponentials a 1000 times
- Let X_i be the outcome for sample i
- Then note that $\mu = E[X_i] = 1/\lambda^2 = 1/0.2^2 = 25$ (Vertical dotted line)
- Let's take n exponentials, take their mean and repeat this 1000 times (Simulation Variances)



Normal Approximation of the Distribution

- Simulate $n = 40$ exponentials a 1000 times
- Let X_i be the outcome for sample i
- Then note that $\mu = E[X_i] = 1/\lambda = 1/0.2 = 5$
- $Var(X_i) = 1/\lambda^2 = 25$
- SE $\sqrt{25/n} = 5/\sqrt{n}$
- Let's take n exponentials, take their mean, subtract off 5, and divide by $5/\sqrt{n}$ and repeat this 1000 times (Simulation Histogram)
- Compare this against normal distribution (Dotted curve)



Sample Mean vs Theoretical Mean

- Mean of simulation means is 4.9681484
- Theoretical mean is 5
- Thus simulation means are a good estimate of the mean and from the first plot we can see that it is an unbiased estimate the mean.

Sample Variance vs Theoretical Variance

- Mean of simulation variances is 25.7760934
- Theoretical variance is `tvar`
- Thus simulation variances are a good estimate of the variance and from the second plot we can see that it is an unbiased estimate of the variance.

Distribution

- Mean of normalized simulation is -0.0243168
- Variance of normalized simulation is 0.906701
- These are close to `0` and `1` respectively which shows that the **Central Limit Theorem** (CLT) is correct for the exponential distribution, i.e. the distribution of averages of iid variables becomes that of the standard normal as the sample size increases.

Appendix

The following shows R code used in the analysis:

- Code that generates the first graph and calculates the simulation and theoretical mean.

```
library(ggplot2)
nosim <- 1000; n <- 40; lambda = 0.2;
tmean <- 1/(lambda);
dmean <- data.frame(
x = c(apply(matrix(rexp(nosim * n, lambda), nosim), 1, mean))
)
ggplot(dmean, aes(x = x, fill = "Simulation Means")) +
geom_density(size = 2, alpha = .2) +
geom_vline(xintercept = tmean, size = 2, linetype = 3) +
theme(legend.position="bottom", legend.title=element_blank());
smean <- mean(dmean$x)
```

- Code that generates the second graph and calculates the simulation and theoretical variance.

```
library(ggplot2)
nosim <- 1000; n <- 40; lambda = 0.2;
tvar <- 1/(lambda^2);
dvar <- data.frame(
x = c(apply(matrix(rexp(nosim * n, lambda), nosim), 1, var))
)
ggplot(dvar, aes(x = x, fill = "Simulation Variances")) +
geom_density(size = 2, alpha = .2) +
geom_vline(xintercept = tvar, size = 2, linetype = 3) +
theme(legend.position="bottom", legend.title=element_blank());
svar <- mean(dvar$x)
```

- Code that generates the third graph and calculates the normalized simulation mean and variance.

```
library(ggplot2)
nosim <- 1000; n <- 40; lambda = 0.2;
cfunc <- function(x, n) sqrt(n) * (mean(x) - 1/lambda) / sqrt(1/lambda^2)
dat <- data.frame(
  x = apply(matrix(rexp(nosim * n, lambda),
                 nosim), 1, cfunc, n)
  )
g <- ggplot(dat, aes(x = x)) +
geom_histogram(alpha = .20,
               binwidth=.3,
               colour = "black",
               fill = "pink",
               aes(y = ..density..))
g <- g + stat_function(fun = dnorm, size = 2, linetype = 3)
g
s2mean <- mean(dat$x)
s2var <- var(dat$x)
```