An investigation on the exponential distribution

Z.C. Neo

Overview

In this report, we will investigate some properties of the distribution of the mean of 40 exponentials. The distribution is obtained from one thousand simulations.

Setup

First we load the required libraries and set some global options.

```
library(knitr)
library(ggplot2)
opts_chunk$set(echo=TRUE, results="asis")
```

We set the seed for reproducibility, define some constants and proceed with the simulation.

```
set.seed(14)
lambda <- 0.2
n <- 40
nSim <- 1000
data <- apply(matrix(data=rexp(n*nSim, lambda), nrow=nSim, ncol=n), 1, mean)
```

Sample statistics versus theoretical statistics

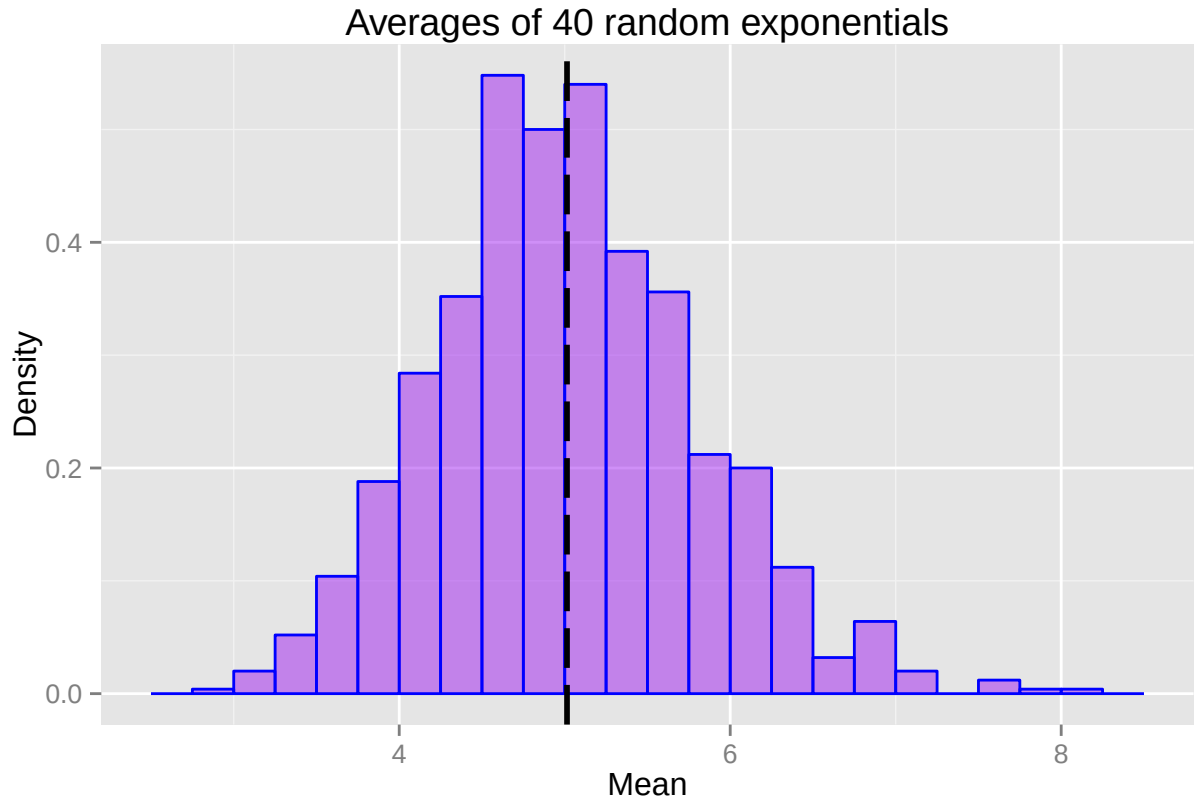
The theoretical mean of the distribution (from 1/lambda) is 5. The sample mean of the distribution is found to be

```
mean(data)
```

[1] 5.013765

As seen from the plot below, the sample mean (dashed line) is very close to the theoretical mean. This can be explained by the Central Limit Theorem (CLT) which states that the distribution of averages of independent and identically distributed (iid) samples is centred at the population mean.

```
plot1 <- ggplot(data.frame(data), aes(x=data))
plot1 + geom_histogram(aes(y=..density..), alpha=0.5, binwidth=0.25, color="blue",
    fill="purple") +
    labs(x="Mean", y="Density", title="Averages of 40 random exponentials") +
    geom_vline(xintercept=mean(data), size=1.0, color ="black", linetype="longdash")
```



The theoretical variance is calculated by

```
(1/lambda)^2 / 40
```

[1] 0.625

The sample variance of the distribution is found to be

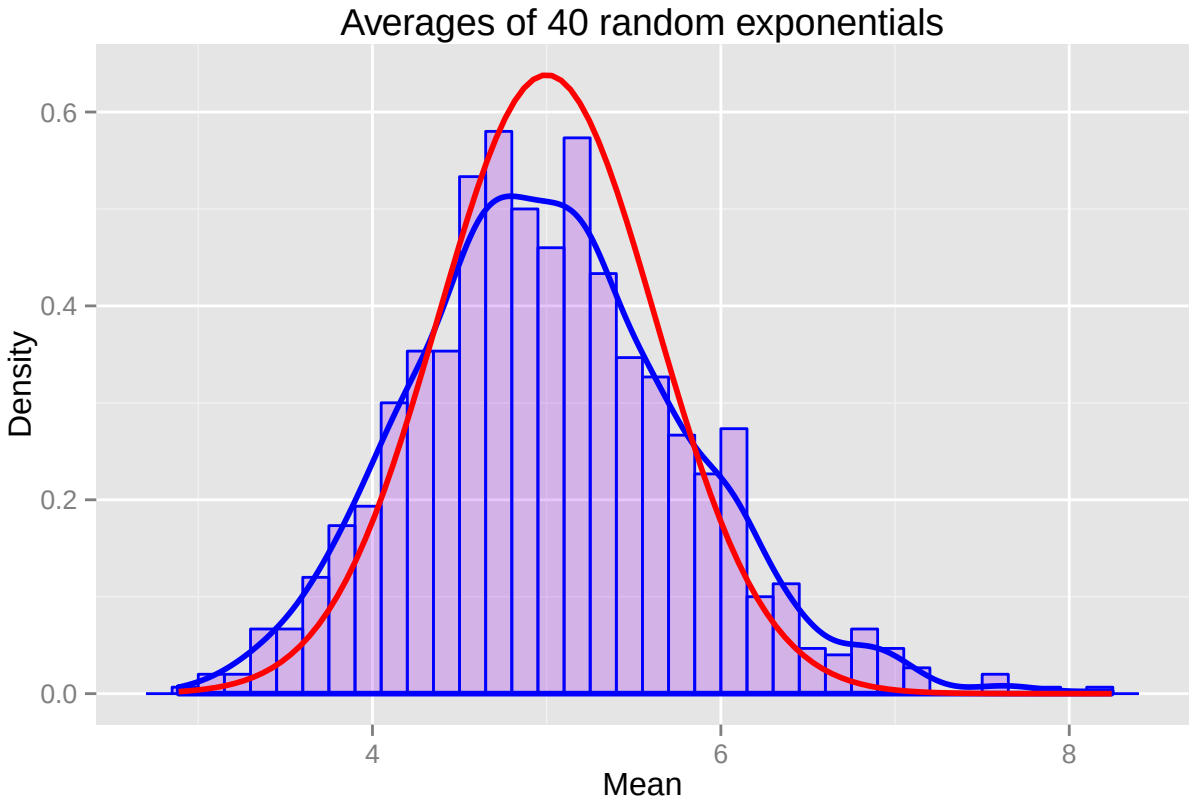
```
var(data)
```

[1] 0.6165999

The two values are close. This can be explained by CLT which states that the distribution of averages of iid samples has standard deviation equal to the standard error of the mean.

Distribution

```
plot2 <- ggplot(data.frame(data), aes(x=data))
plot2 + geom_histogram(aes(y=..density..), alpha=0.25, binwidth=0.15, color="blue",
    fill="purple") +
    labs(x="Mean", y="Density", title="Averages of 40 random exponentials") +
    geom_density(size=1.0, color="blue") +
    stat_function(fun=dnorm, args=list(mean=5, sd=((1/lambda)^2 / 40)), size=1.0,
                  color="red")
```



The blue line is the empirical density while the red line is the theoretical density. This shows that distribution is approximately normal.