Examinging sample and theoretical distributions

MazinMelb

22 Nov 2015

Overview

This project is considering the sample and theoretical distribution of randomly generated exponentials for the Coursera subject Statistical Inference. We'll be looking at the distribution of sample means to consider how the sample means compare to the theoretical population mean in accordance with the Central Limit Theorem.

Simulations:

We know that the mean of an exponential distribution is 1/lambda and the standard deviation is also 1/lambda. This project has simulated 1000 samples of 40 exponentials with lambda set to 0.2.

The sample means were generated using the mean() function, and the sample variances were generated using the var() function on the exact same data:

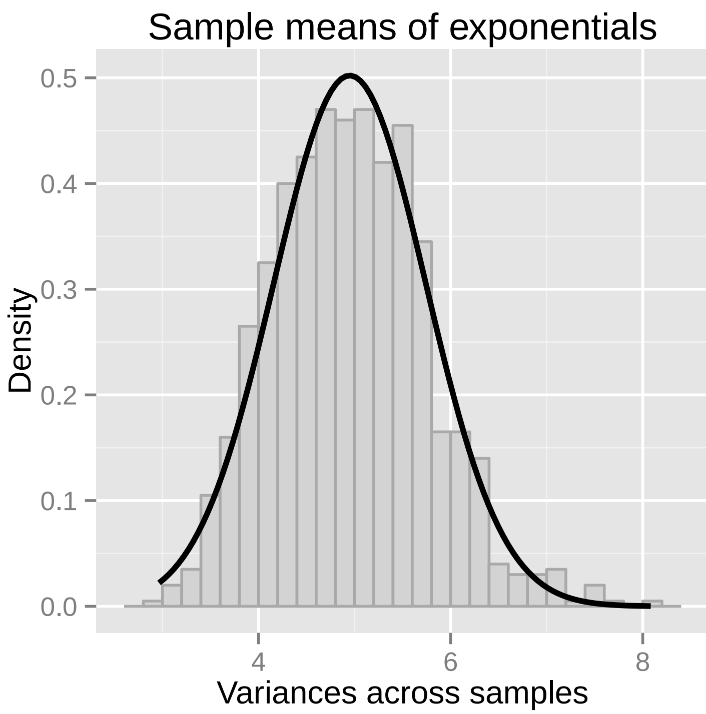
```
mns = NULL
vars = NULL
for (i in 1 : 1000) {
  sample <- rexp(40,0.2)
  mns = c(mns, mean(sample))
  vars = c(vars,var(sample))
}
```

All of the sample means were represented as standardised scores, so that each mean is represented by it's distance in standard deviations from the mean of means:

```
zs = NULL
zs = sapply(mns,function(x) (x-mean(mns))/sd(mns))
```

Sample Mean versus Theoretical Mean:

The sample mean for each of the 1000 samples of the 40 random exponentials can be visualised as:



The graph helps visualise the variablity in the mean of 40 random variables across the samples.

As predicted by the CLT graphing the mean of 1000 samples of 40 random exponential numbers gives an almost normal distribution centred around the theoretical mean of 5 (1/lambda).

Correspondingly, the mean of the sample means gives the result:

```
mean(mns)
```

```
## [1] 4.949416
```

This result can be seen to be very close to the thoeretical population mean of 5.

Sample Variance versus Theoretical Variance:

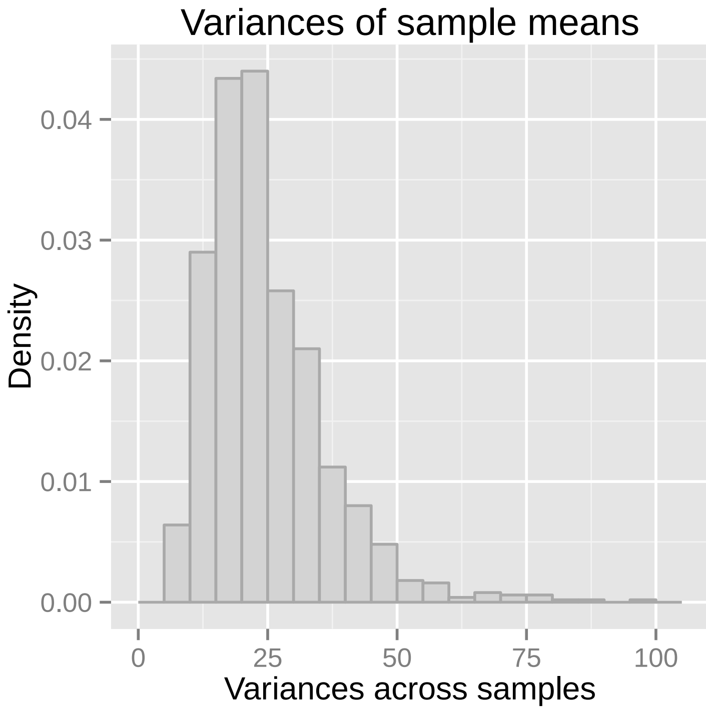
The theoretical variance of the population of exponentials is 1/lambda^2. When lambda is 0.2, as for our samples, then the theoretical variance is 25.

The mean of the sample variances gives a result close to the theoretical population mean:

```
mean(vars)
```

```
## [1] 24.69852
```

The variances in our sample means can be visualised by plotting the variances of each sample, these variances are centred at the mean sample variance.



Distribution:

Representing the sample means as standardised scores, that is, representing them by their distance from the mean of means in standard deviations, helps compare the distribution of sample data to the normal distribution. We can see the variances of the sample means is a nearly normal distribution. We can see that most of the data falls within the expected 1.96 standard deviations from the mean.