Statistical Inference Simulation Project

Anton Hesse

9/10/2020

Exponential Simulation

We will simulate random exponential distributions in R and compare our results to the central limit theorum. We begin by loading the ggplot2 package.

```
library(ggplot2)
```

Next we will define our variables. We will have distributions of averages of 40 random exponentials. Our rate parameter, lambda, will be set to 0.2 and we will repeat this simulation 1000 times.

```
N <- 40 # number of averages
lambda <- 0.2 # rate parameter
nSims <- 1000
```

Next we will simulate our data with the replicate function.

```
set.seed(5730)
simData <- replicate(nSims, rexp(n = N, rate = lambda))
head(simData[,1:5])
```

```
##            [,1]      [,2]      [,3]     [,4]      [,5]
## [1,]  2.8351482 9.0742983 7.7051910 3.880677 2.3540790
## [2,] 16.2623219 3.9763289 0.4628226 2.015746 1.8420522
## [3,]  9.4336542 3.7246719 9.4606501 7.634500 3.4683237
## [4,]  0.2652534 6.2828984 0.5259829 6.131186 0.4173234
## [5,]  5.0284578 0.9849492 1.3245374 7.120590 7.3106147
## [6,] 14.6610194 2.0557058 1.2355581 1.986237 4.8032901
```

We we now calculate the mean and variance of each of the 1000 replicates of 40 samples.

```
means <- apply(simData, MARGIN = 2, FUN = mean)
head(means)
```

```
## [1] 4.967865 4.904491 3.441910 3.592864 4.293440 5.499593
```

```
vars <- apply(simData, MARGIN = 2, FUN = var)
head(vars)
```

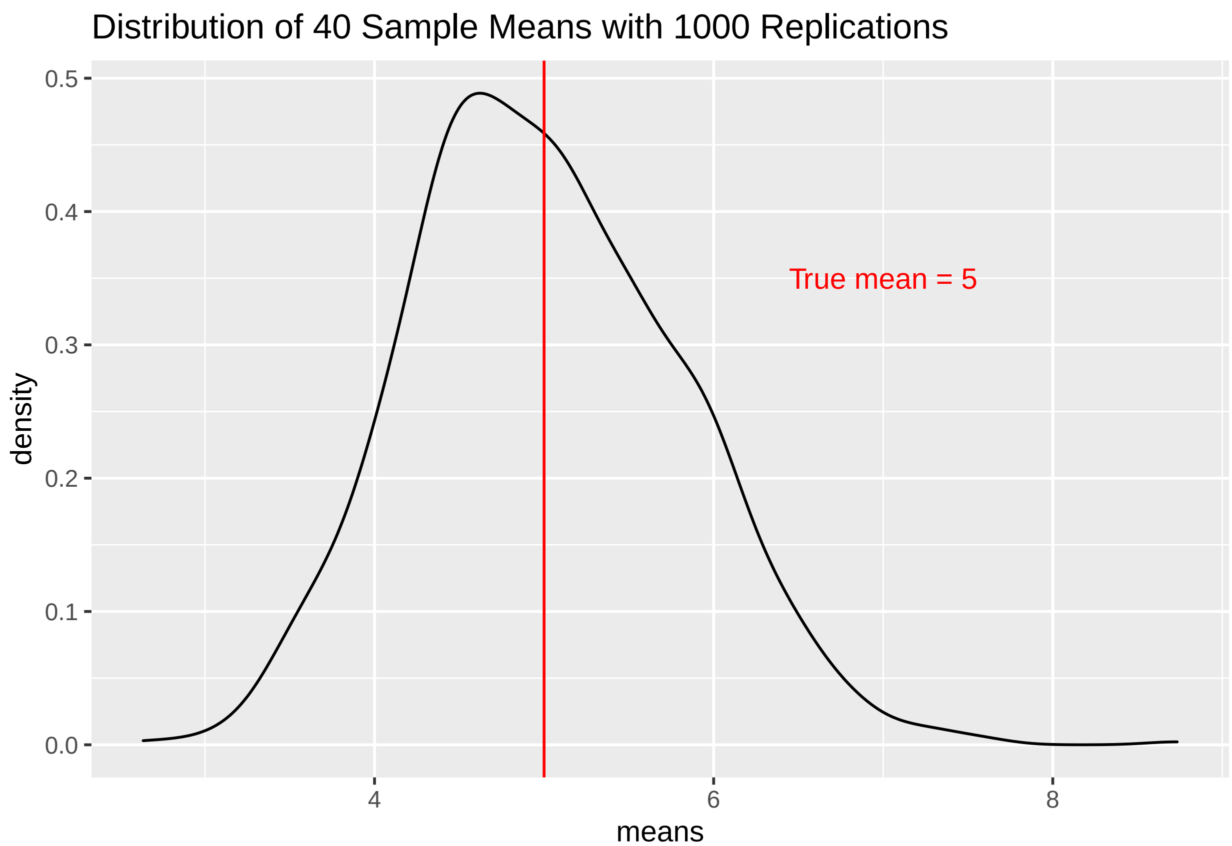
```
## [1] 19.548645 19.563122 15.073803  9.748262 13.974489 27.432585
```

Prior to plotting our results, we will calculate the true mean and variance for our chosen value of lambda.

```
truMean <- 1/lambda
truSD <- 1/lambda
truVar <- truSD^2
```

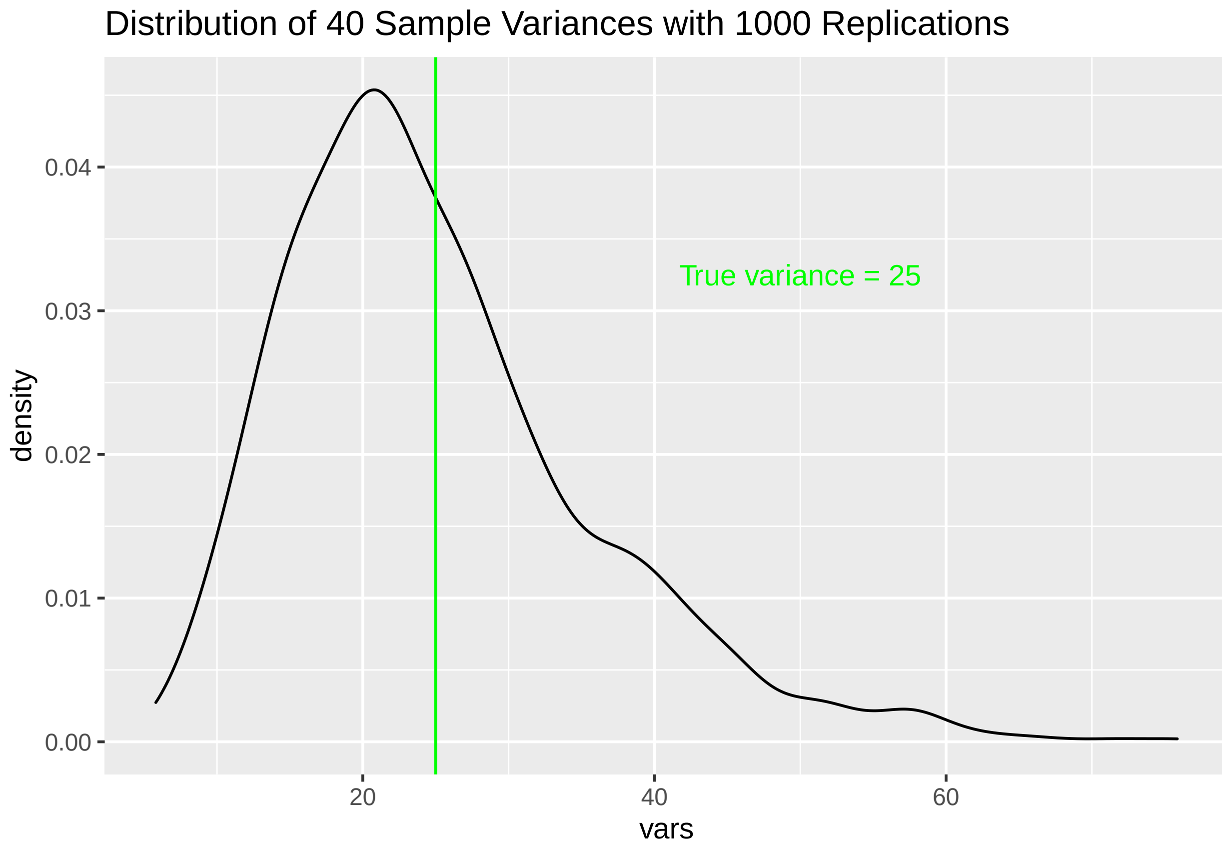
We can now compare our sampling distributions of the means and variance with the true values. We will create density plots and add a vertical line that indicates the true mean and variance values.

```
meanDist <- ggplot(data = as.data.frame(means)) +
    geom_density(aes(x = means)) +
    geom_vline(xintercept = truMean, color = "red") +
    ggtitle("Distribution of 40 Sample Means with 1000 Replications") +
    annotate("text", x = 7, y = 0.35, label = "True mean = 5", color = "red")
meanDist
```



The mean of the mean sampling distribution for our simulation of 5.004 was very similar to the true mean value of 5. This density plot has a characteristic bell shape.

```
varDist <- ggplot(data = as.data.frame(vars)) +
    geom_density(aes(x = vars)) +
    geom_vline(xintercept = truVar, color = "green") +
    ggtitle("Distribution of 40 Sample Variances with 1000 Replications") +
    annotate("text", x = 50, y = 0.0325, label = "True variance = 25", color = "green")
varDist
```



The mean variance of the variance sampling distribution simulation of 25.064 was very similar to the true mean variance of 25. This density plot has a characteristic bell shape. However, in comparison to the mean density plot, this is somewhat right skewed given that these values are squares. Although this skewness makes it somewhat less bell shaped than the means plot, it still is bell shaped.