Statistical Inference Project

Mike Taylor

13 July 2015

A simulation exercise investigating the exponential distribution

In this analysis multiple random samples will be taken from an exponential distribution. The mean of each of these samples will be stored in a new variable, and the mean and variance of this variable will be compared to the original population (the exponential random variable).

Simulations

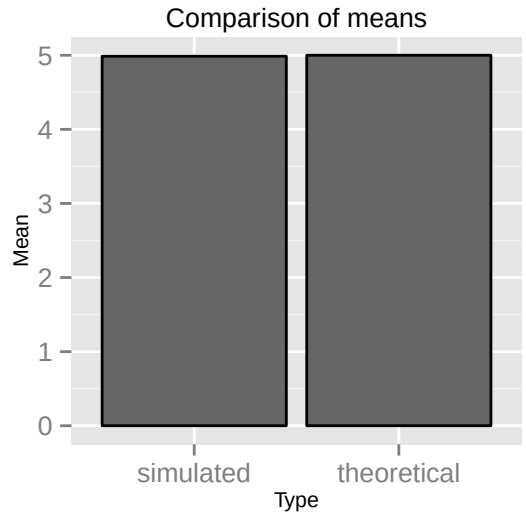
```
#First set the random number seed to make the results reproducible
set.seed(11)

#create a variable to store the results in
means = numeric()

#run 1000 simulations
for (i in 1:1000) {
  #store the mean of 40 exponentials to the means variable, using the rate 0.2
  means[i] <- mean(rexp(40, 0.2))
}
```

Sample Mean versus Theoretical Mean

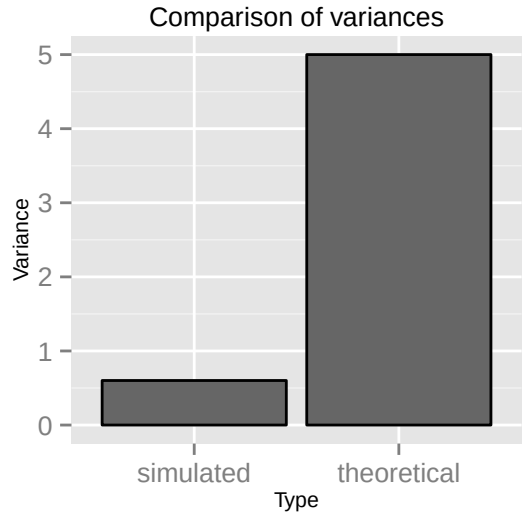
```
library(ggplot2)
ggplot(data.frame(Mean = c(5, mean(means)), Type = c("theoretical", "simulated")),
       aes(x = Type, y = Mean)) +
  geom_bar(fill = "grey40", colour = "black", stat = "identity") +
  ggtitle("Comparison of means") +
  theme(plot.title = element_text(size = 10),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8))
```



The figure above shows the mean of the distribution calculated from the simulations alongside the theoretical mean. The mean derived from simulation provides a very good (very close) estimate of the distribution mean. The mean of the distribution of sample means is 4.9871567.

Sample Variance versus Theoretical Variance

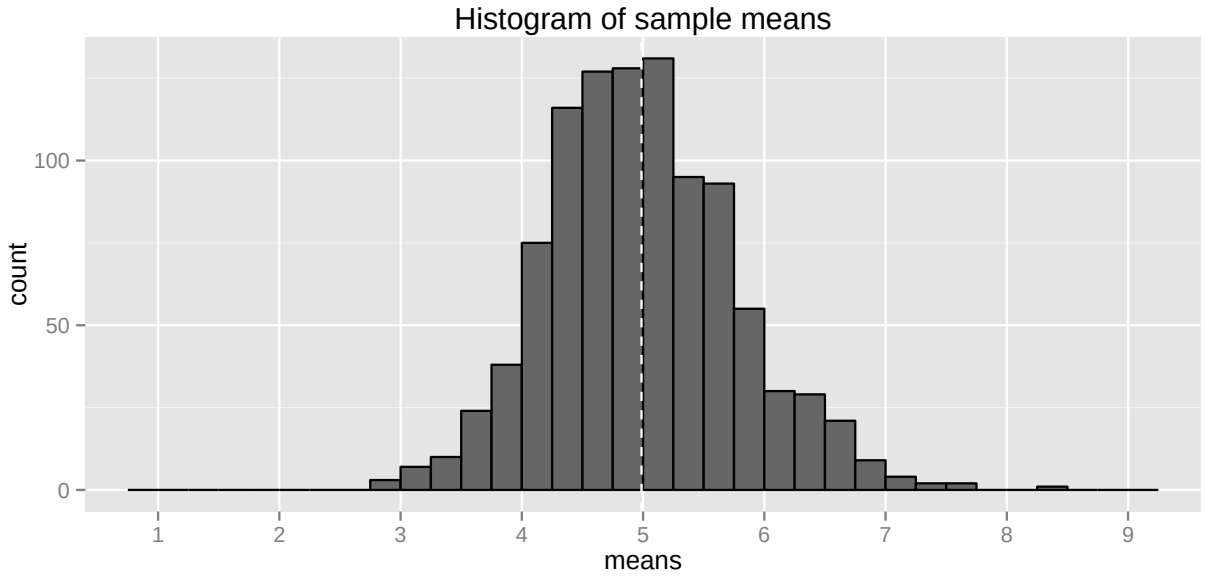
```
ggplot(data.frame(Variance = c(5, sd(means) ** 2), Type = c("theoretical", "simulated")),
       aes(x = Type, y = Variance)) +
  geom_bar(fill = "grey40", colour = "black", stat = "identity") +
  ggtitle("Comparison of variances") +
  theme(plot.title = element_text(size = 10),
        axis.title.x = element_text(size = 8),
        axis.title.y = element_text(size = 8))
```



The figure above shows the variance of the sample means alongside the original distribution variance. The variance is much smaller than the original distribution variance. The variance of the sample means is 0.6009383.

Test whether distribution of sample means is approximately normal

```
ggplot(data.frame(means), aes(x = means)) +
  geom_histogram(fill = "grey40", binwidth = 0.25, colour = "black") +
  scale_x_discrete(limits = 1:9) +
  geom_vline(xintercept = mean(means), linetype = "dashed", color = "white") +
  ggtitle("Histogram of sample means")
```



The above figure shoes a histogram of the sample means, with a line showing the mean of these samples. As you can see it has a clear peak centered around the mean and looks both symmetrical and bell shaped.

Also going to have a look at how many of the sample means are within 1, 2 or 3 standard deviations of the mean.

Within 1: 0.708 Within 2: 0.948 Within 3: 0.996

As seen above, the proportion of the distribution of sample means within 1, 2 or 3 standard deviations is very close to the proportions seen for a true normal distribution (0.68, 0.95 and 0.997 respectively).