Inferential Statistics Assignment 1

Karl Ranson

31 August 2016

Introduction

This PDF meets the requirements of Inferential Statistics assignment part 1, which are:

(1) Perform exporatory analysis on exponential data to demonstrate some basic characteristics;
(2) Compare the sampled vs theoretical mean; and
(3) Compare the sampled vs theoretical variance.

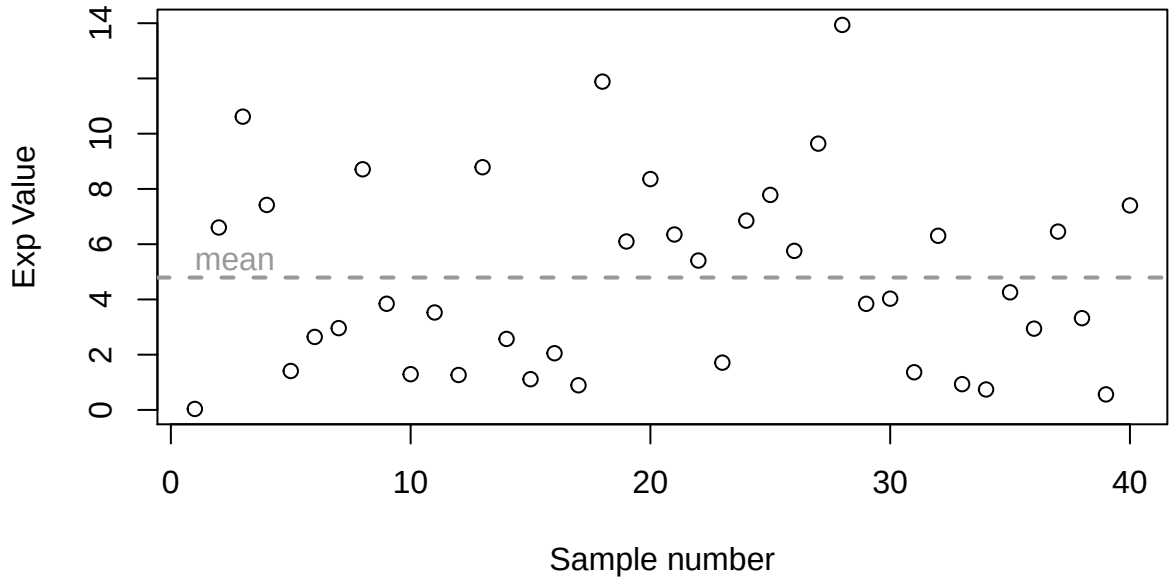
To keep within the 3 page report limit, all R code is in Appendix 2.

Part 1.1 - Simulations

10,000 sample sets of 40 exponential randoms with lambda = 0.2 were created.

The below plot shows the first set of 40 random samples.

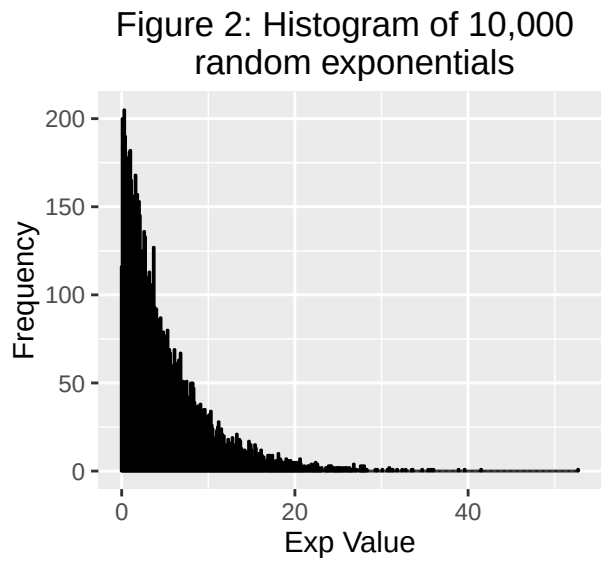
Figure 1: First set of 40 exp samples



Part 1.2 - Means

The 'apply' function was used to create a 1 dimensional matrix of the 10,000 means of 40 samples.

```
## Warning: The plyr::rename operation has created duplicates for the
## following name(s): (`linetype`)
```



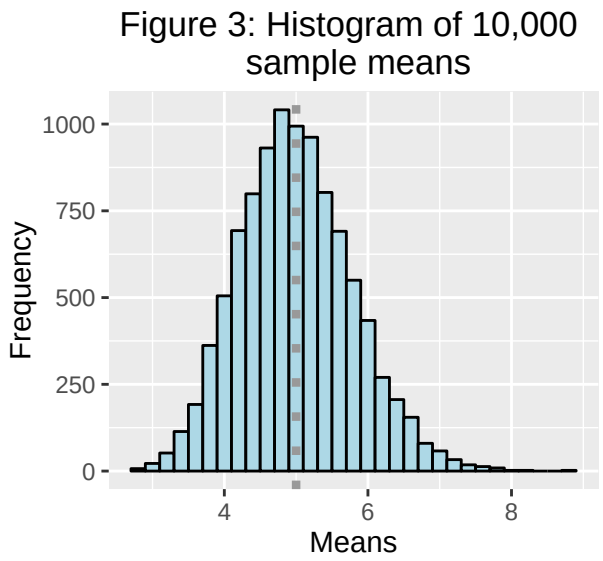
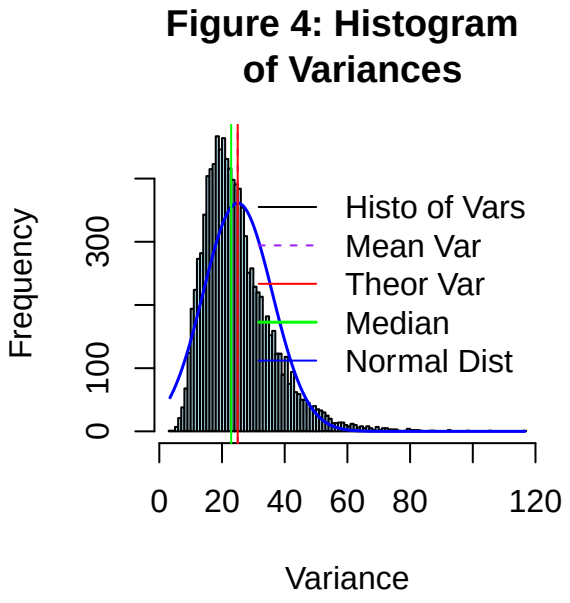


Figure 2 shows a histogram of 10,000 random exponentials. You can see it closely resembles an exponential distribution.

Figure 3 is a the histogram of the means and exhibits a normal distribution shape. The theoretical mean, 1/lambda = 1/(0.2)=5, is shown by the dotted line. You can see it matches quite well with middle of the histogram.

Part 1.3 - Variance Histogram

Similar to the above, instead of the means the variance of 10,000 sample sets was taken and represented in a histogram in



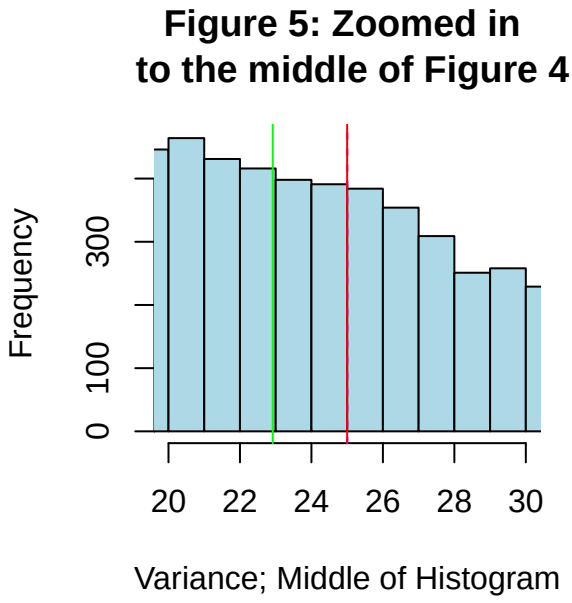


Figure 4 shows the histogram of variances, with a normal distribution overlaid for comparison. The normal curve has the same mean and variance as the sample variances. You can see that the variance histogram does not exhibit as normal behaviour as the means histogram (Figure 3); with a steeper left hand side and a fatter right hand side with longer tail.

Figure 5 is a zoom in on the middle bin of the Fig 4 plot, clearly showing the theoretical variance (in red) and the mean of the sample variances (in purple, barely visible under the red), are almost identical. The median is also shown in green. The difference between the median and mean is the result of the long right hand tail.

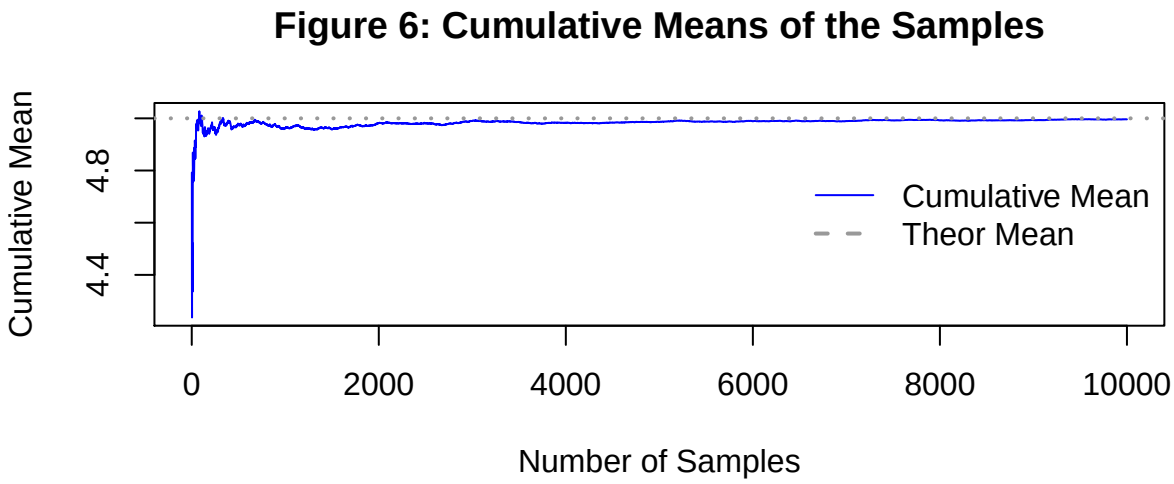


Figure 6 shows how the cumulative mean as the number of samples increases quickly converges on the theoretical mean, in grey. This is consistent with the law of large numbers.

Conclusions

Figure 6 shows the law of large numbers in effect, as the cumulative mean of the samples converge with the theoretical mean shown in grey.

The mean histogram (Figure 3) exhibits normal distribution tendencies, and thus is an example of the Central Limit Theorem; where under certain conditions the means of independent variables tend to have normal distributions. This theorem holds true even if the underlying population is not normal; which is this case.

Appendix: All R code

Libraries

```
knitr::opts_chunk$set(cache=TRUE)
library(grid)
library(ggplot2)
library(gridExtra)
library(moments)
library(stats)
```

Part 1.1: Exp samples code - simulations

```
plot(raw_rand_set[1,],main = "Figure 1: First set of 40 exp samples", xlab = "Sample number",ylab = "Exp
abline(h = mean(raw_rand_set[1,]),col = "gray60", lwd=2, lty=2)
text(1,mean(raw_rand_set[1,])+.2, "mean", col = "gray60", adj = c(0, -.1))
```

```
n <- 40
lambda <- 0.2
p <- rexp(n,lambda)
raw_rand_set <- matrix(rexp(n * 10000, lambda), ncol = n)
```

Part 1.2: Exp samples code - means

```
par(mfrow = c(1,2))
varc <- apply(raw_rand_set,1,var)

MaxFreq <- max(hist(varc,plot="FALSE")$counts)
breaks=100
h1<-hist(varc, breaks=breaks, col="light blue", xlab="Variance",
    main="Figure 4: Histogram \n of Variances") # Add a Normal Curve
xfit<-seq(min(varc),max(varc),length=length(varc*1))
yfit<-dnorm(xfit,mean=mean(varc),sd=sd(varc))
yfit <- yfit*diff(h1$mids[1:2])*length(varc)
lines(xfit, yfit, col="blue", lwd=1.5)
abline(v = mean(varc), lty = 2,col="purple")
abline(v = 1/lambda^2, lty = 1,col="red")
abline(v = median(varc), lty = 1,col="green")

legend("right", bty = "n", xjust = 1, yjust = 1, col = c("black", "purple", "red","green","blue"), lwd =

h2<-hist(varc, breaks=breaks, col="light blue", xlab="Variance; Middle of Histogram", xlim = c(mean(varc
    main="Figure 5: Zoomed in \n to the middle of Figure 4")
abline(v = mean(varc), lty = 2,col="purple")
abline(v = 1/lambda^2, lty = 1,col="red")
abline(v = median(varc), lty = 1,col="green")
```

Part 1.3: Exp samples code - vars

```
par(mfrow = c(1,2))
varc <- apply(raw_rand_set,1,var)

MaxFreq <- max(hist(varc,plot="FALSE")$counts)
breaks=100
h1<-hist(varc, breaks=breaks, col="light blue", xlab="Variance",
    main="Figure 4: Histogram \n of Variances") # Add a Normal Curve
xfit<-seq(min(varc),max(varc),length=length(varc*1))
yfit<-dnorm(xfit,mean=mean(varc),sd=sd(varc))
yfit <- yfit*diff(h1$mids[1:2])*length(varc)
lines(xfit, yfit, col="blue", lwd=1.5)
abline(v = mean(varc), lty = 2,col="purple")
abline(v = 1/lambda^2, lty = 1,col="red")
abline(v = median(varc), lty = 1,col="green")

legend("right", bty = "n", xjust = 1, yjust = 1, col = c("black", "purple", "red","green","blue"), lwd =

h2<-hist(varc, breaks=breaks, col="light blue", xlab="Variance; Middle of Histogram", xlim = c(mean(var
    main="Figure 5: Zoomed in \n to the middle of Figure 4")
abline(v = mean(varc), lty = 2,col="purple")
abline(v = 1/lambda^2, lty = 1,col="red")
abline(v = median(varc), lty = 1,col="green")
```

Part 1.3: Exp samples code - cumulative means

```
cumulative_means <- cumsum(means)/(1:length(means))
plot(cumulative_means, type="l", col="blue",main = "Figure 6: Cumulative Means of the Samples", xlab =
abline(h = 5, col = "grey60", lty = 3,lwd=2 )
legend("right", bty = "n", xjust = 1, yjust = 1, col = c("blue", "grey60"), lwd = c(1,2), lty = c(1,2),
```