Statistical Inference Course Project Part 1

Joe Schaal

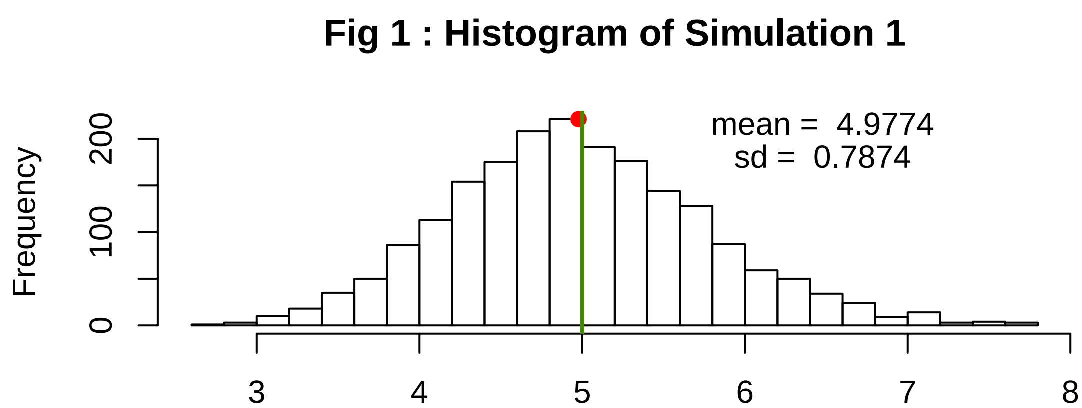
Thursday, November 06, 2014

Simulation of the Averages of Exponential Random Variables

This report discuss the results of a simulation of the **mean** value of **40** I.I.D. exponential random variables with lambda λ = **0.2**. The simulation compares the distribution of **2000** trials with that of a normal distribution. Mathematically, the **mean** u and **standard deviation** s of an exponential distribution with rate λ is $1/\lambda$. For this simulation, u and s are each equal to **5**.

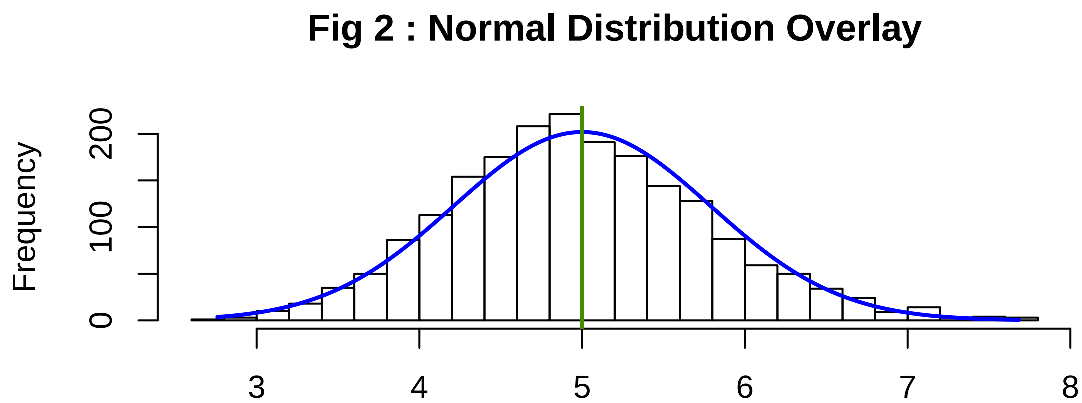
Generating 2000 trials results in the following histogram (Fig 1):

```
sim1 <- apply(matrix(rexp(n*trials,lambda),trials),1,mean)
```



The **theorectical mean** of the distribution (represented by the green line in the chart) is **5** which compares favorably to the **simulation mean** of **4.9774** (represented by the red point in the chart) and the distribution is centered near its theoretical mean. The theoretical variance of a sample mean for a sample of size 40 (represented by its **standard deviaition** s) is $\frac{\sigma}{\sqrt{n}} = \frac{(1/\lambda)}{\sqrt{40}} = \frac{5}{6.3246}$ =**0.7906**. This also compares favorably to the **actual standard deviation** of **0.7874**.

If we overlay a normal distribution (Fig 2), with the theoretical **mean** 5 (represented by the green line) and **standard deviation** 0.7906, we see that normal distribution is very similar to the simulation distribution.

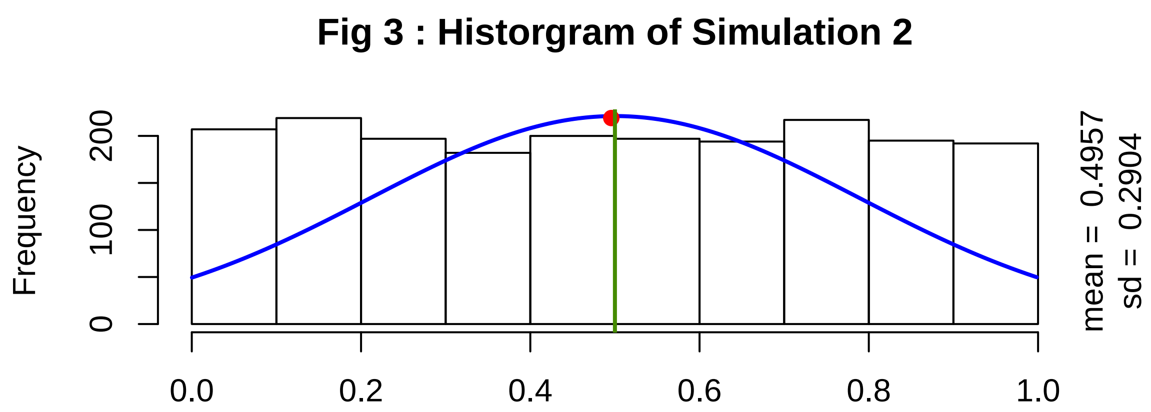


This exercise shows that in general, when estimating the averages of a series of **I.I.D.** random variables from a common distribution, for sufficiently large n, the distribution is centered around the **expected value** of the population average, has a **variance** that is proportional to the variance of the population: $\frac{\sigma^2}{n}$, and approaches the distribution of a normal distribution with the equivalent mean and standard deviation.

Comparison of Individual to Average of Individual Uniform Random Variables

This differs from a comparison against a sample of individual random variables from a population, which accordingly, will follow its inherent distribution. For example, for a standard uniform variable with range (0,1), the mean u is $\frac{1}{2}(0+1) = \mathbf{0.5}$ (represented by the green line in Fig 3) and standard deviation $s = \sqrt{\sigma^2} = \sqrt{\frac{1}{12}(1-0)^2}$ =**0.2887**. Simulating **2000** of individual uniform distributions results in the following distribution (Fig 3):

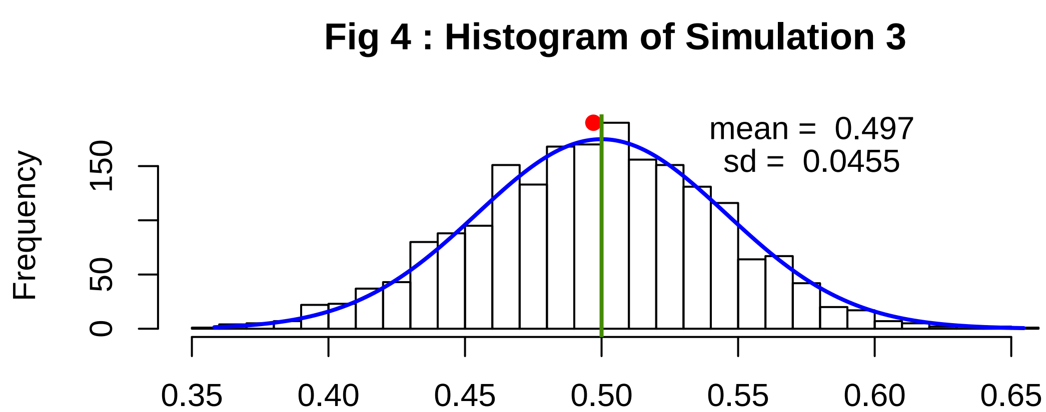
```
sim2 <- runif(trials)
```



The **mean** $u = 0.4957$ (represented by the red point) and **standard deviation** $s = 0.2904$ of the sample are very close to that of the underlying distribution but the distribution does not follow that of a normal distribution.

However, if we instead sample the **averages** of a series of uniform random variables, such as sampling the average of $n = 40$ random variables, we see a very different pattern. For the sample of the mean of the population of uniform variables, we expect the mean to approach $\frac{1}{2}(a+b) = 0.5$ (represented by the green line in Fig 4) and the standard deviation to approach $s = \sqrt{\frac{\sigma^2}{n}} = \sqrt{\frac{(1/12)(1-0)^2}{40}} = 0.0456$.

```
sim3 <- apply(matrix(runif(n*trials),trials),1,mean)
```



In the simulation example, the **mean** of the simulation is **0.497** (represented by the red point) and the **standard deviation** is **0.0455**. This compares favorably with that of the theoretical mean and standard distribution, as well as that of a normal distribution with mean **0.5** and standard deviation **0.0456**.

R Code for Included Figures

Initialization Code:

```
set.seed(5555)
n <- 40
trials <-2000
lambda <- 0.2
expMean <- 1/lambda
expSD <- (1/lambda)/sqrt(n)
figureCount <- 1

figureNumber <- function(name,figureCount) {
    figureText <- paste("Fig",as.character(figureCount),":",name)
    figureCount <<- figureCount+1
    return(figureText)
}
```

Figure 1 Code

```
sim1 <- apply(matrix(rexp(n*trials,lambda),trials),1,mean)
simMean <- mean(sim1)
simSD <- sd(sim1)
simHist <- hist(sim1,breaks="Scott",
                main = figureNumber("Histogram of Simulation 1",figureCount),
                xlab="")
yMax <- max(simHist$counts)
meanText <- paste("mean = ",as.character(round(simMean,4)))
sdText <- paste("sd = ",as.character(round(simSD,4)))
points(simMean,yMax,col="red",cex=1,pch=21,bg="red")
text(x = simMean+1.5 ,y = yMax-5, meanText)
text(x = simMean+1.5 ,y = yMax-40, sdText)
abline(v=expMean,col="chartreuse4",lwd=2)
```

Figure 2 Code

```
simHist <- hist(sim1,breaks="Scott",
                main = figureNumber("Normal Distribution Overlay",figureCount),
                xlab = "")
xfit<-seq(min(sim1),max(sim1),length=1000)
yfit<-dnorm(xfit,mean=expMean,sd=expSD)
yfit <- yfit*diff(simHist$mids[1:2])*length(sim1)
lines(xfit, yfit, col="blue", lwd=2)
abline(v=expMean,col="chartreuse4",lwd=2)
```

Uniform Distribtution Initialization

```
a <- 0
b <- 1
unfMean <- 1/2*(a+b)
unfSD <- sqrt((1/12)*(b-a)^2)
unfSampleSD <- sqrt((unfSD^2)/n)
```

Figure 3 Code

```
sim2 <- runif(trials)
simHist <- hist(sim2,breaks="Scott",
                main =  figureNumber("Historgram of Simulation 2",figureCount),
                xlab="")
simMean <- mean(sim2)
simSD <- sd(sim2)
xfit<-seq(min(sim2),max(sim2),length=1000)
yfit<-dnorm(xfit,mean=unfMean,sd=unfSD)
yfit <- yfit*diff(simHist$mids[1:2])*length(sim2)
lines(xfit, yfit*.80, col="blue", lwd=2)
yMax <- max(simHist$counts)
meanText <- paste("mean = ",as.character(round(simMean,4)))
sdText <- paste("sd = ",as.character(round(simSD,4)))
points(simMean,yMax,col="red",cex=1,pch=21,bg="red")
mtext(text = meanText,side = 4,line=0)
mtext(sdText,4,line=1)
abline(v=unfMean,col="chartreuse4",lwd=2)
```

Figure 4 Code

```
sim3 <- apply(matrix(runif(n*trials),trials),1,mean)
simHist <- hist(sim3,breaks="Scott",
                main = figureNumber("Histogram of Simulation 3",figureCount),
                xlab="")
simMean <- mean(sim3)
simSD <- sd(sim3)
xfit<-seq(min(sim3),max(sim3),length=1000)
yfit<-dnorm(xfit,mean=unfMean,sd=unfSampleSD)
yfit <- yfit*diff(simHist$mids[1:2])*length(sim3)
lines(xfit, yfit, col="blue", lwd=2)
yMax <- max(simHist$counts)
meanText <- paste("mean = ",as.character(round(simMean,4)))
sdText <- paste("sd = ",as.character(round(simSD,4)))
points(simMean,yMax,col="red",cex=1,pch=21,bg="red")
text(x = simMean+.08 ,y = yMax-5, meanText)
text(x = simMean+.08 ,y = yMax-35, sdText)
abline(v=unfMean,col="chartreuse4",lwd=2)
```