P_M6_1

September 1, 2022

1 Module 6 Peer Review Assignment

2 Problem 1

Suppose X and Y are independent normal random variables with the same mean μ and the same variance σ^2. Do the random variables $W = X + Y$ and $U = 2X$ have the same distribution?
Explain.

No

Explain.: consider tht X and Y are independent ,E(W)=E(X+Y)=E(X)+E(Y)=2μ=*E(U),*

V(w)=V(X+Y)=V(X)+V(Y)=$\boldsymbol{2\sigma^2}$

While V(U)=V(2X)=4V(X)=4σ^2,so W and U do not have the same distribution

3 Problem 2: Central Limit Theorem and Simulation

a) For this problem, we will be sampling from the Uniform distribution with bounds [0, 100]. Before we simulate anything, let's make sure we understand what values to expect. If $X \sim U(0,100)$, what is $E[X]$ and $Var(X)$?

E(x)=$\frac{0+100}{2} = 50$ $\quad v_{ar}(X) = \frac{(100-0)^2}{12} = \frac{2500}{3}$

b) In real life, if we want to estimate the mean of a population, we have to draw a sample from that population and compute the sample mean. The important questions we have to ask are things like:

- Is the sample mean a good approximation of the population mean?
- How large does my sample need to be in order for the sample mean to well-approximate the population mean?

Complete the following function to sample n rows from the $U(0,100)$ distribution and return the sample mean. Start with a sample size of 10 and draw a sample mean from your function. Is the estimated mean a good approximation for the population mean we computed above? What if you increase the sample size?

[13]:
```
library(purrr)
flag=vector()
flag1=vector()
r<-vector()
uniform.sample.mean = function(n){
    r<-runif(n,0,100)
    sample.mean<-mean(r)
```

```
    return(sample.mean)
}
 x<-rep(append(flag,10),5)
 x1<-rep(append(flag1,1000),5)
 res10=map(x,uniform.sample.mean) #repeat uniform.sample.mean(10) for 5 times
 res1000=map(x1,uniform.sample.mean) #repeat uniform.sample.mean(1000) for 5 
 ↪times
 res10
 res1000
```

1. 43.0399694480002
2. 48.9053044631146
3. 40.1497644395567
4. 43.1483368319459
5. 54.6743633621372

1. 50.3318975010421
2. 51.2685448225122
3. 3. 49.871335292561
4. 50.4156037507579
5. 50.2558628635248

We notice that when the size is equal to 10 ,the sim mean is not a good approximation for the population mean we computed. When we increase the size from 10 to 1000,the simulated mean becomes a good approximation of the mean we computed

c) Notice, for a sample size of n, our function is returning an estimator of the form

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

That means, if each X_i is a random variable, then our sample mean is also a random variable with its own distribution. We call this disribution the sample distribution. Let's take a look at what this distribution looks like.

Using the `uniform.sample.mean` function, simulate $m = 1000$ sample means, each from a sample of size $n = 10$. Create a histogram of these sample means. Then increase the value of n and plot the histogram of those sample means. What do you notice about the distribution of $\bar{X}$? What is the mean μ and variance σ^2 of the sample distribution?

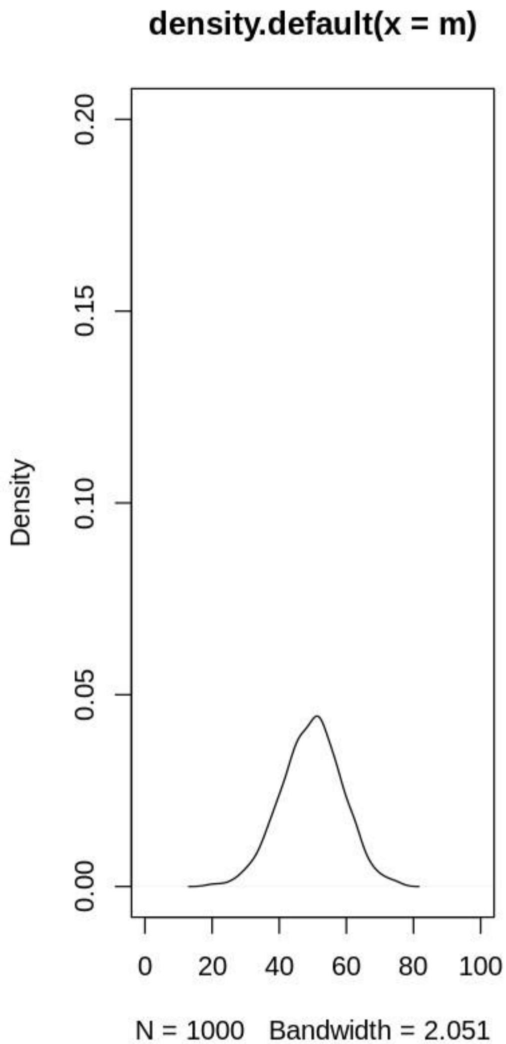
[49]:
```
m=vector()
n=vector()
x=1
repeat{
    u<-uniform.sample.mean(10)
    m<-c(m,u)
    x=x+1
    if(x>1000){break}
}
x=1
repeat{
    u<-uniform.sample.mean(100)
    n<-c(n,u)
    x=x+1
    if(x>1000){break}
}
par(mfrow=c(1,2))
plot(density(m),xlim=c(0,100),ylim=c(0,0.2)) #size =10 , on the left side
plot(density(n),xlim=c(0,100),ylim=c(0,0.2)) #size =100 ,on the right side
var(m)
mean(m)
var(n)
mean(n)
```

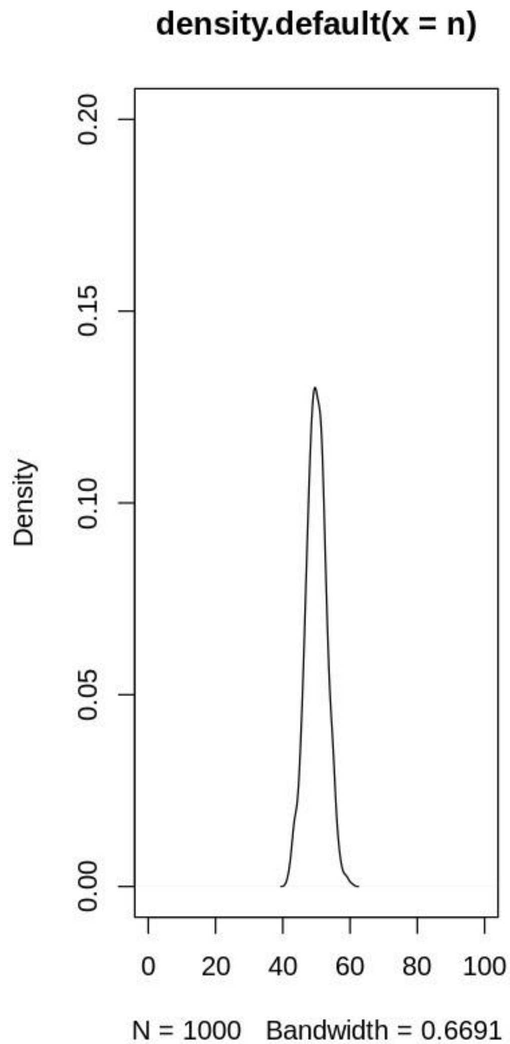
82.3284887749428

49.699511923315

9.00375947570503

49.9548567096433



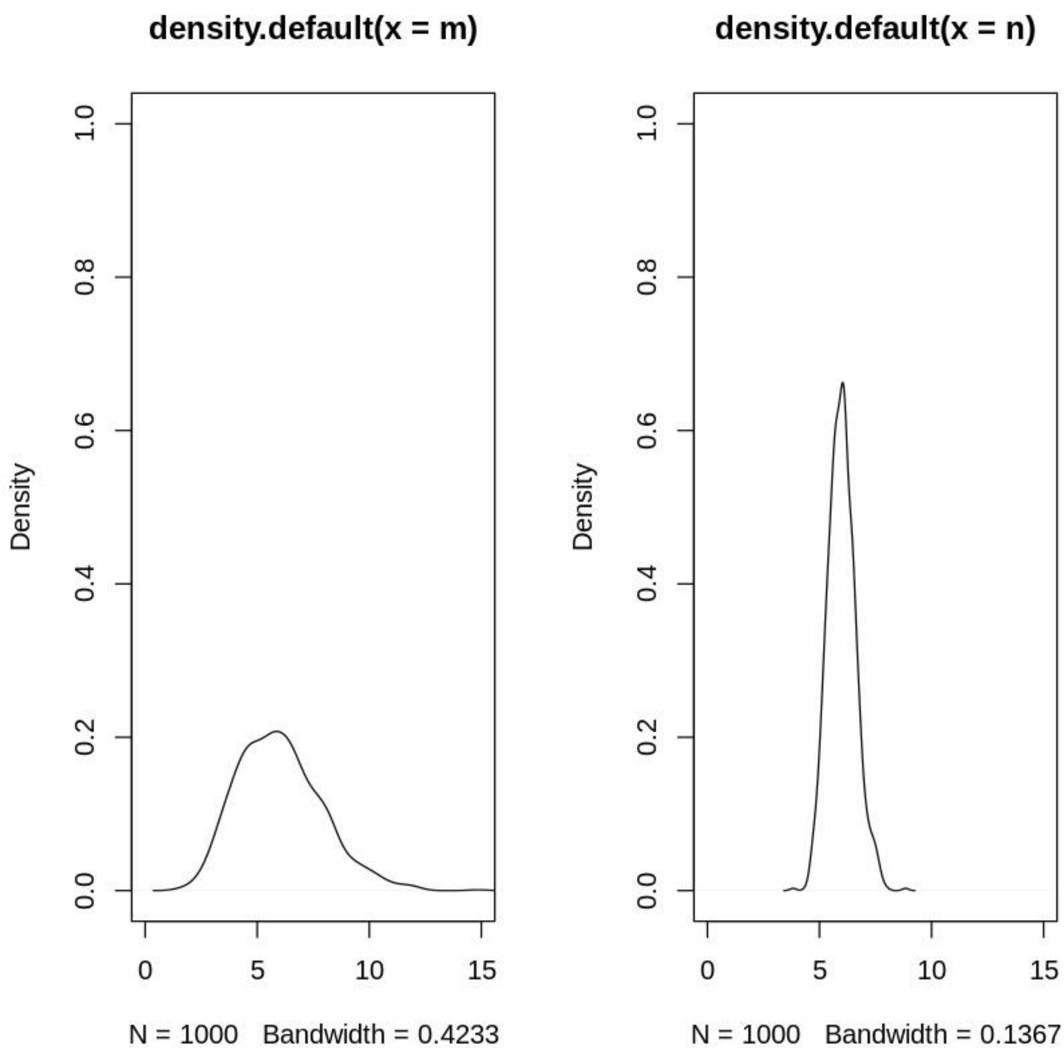


The distribution looks like a normal distribution ,when we increase the size, the variance of the distribution decreases .the mean and variance are shown between the curves and the code block 49.

d) Recall that our underlying population distribution is $U(0,100)$. Try changing the underlying distribution (For example a binomial(10, 0.5)) and check the sample distribution. Be sure to explain what you notice.

[38]:
```
n=vector()
m=vector()
exp.mean = function(n){
    r<-rexp(n,1/6)
    e<-mean(r)
    return(e)
}
x=1
```

```
repeat{
    u<-exp.mean(10)
    m<-c(m,u)
    x=x+1
    if(x>1000){break}
}
x=1
repeat{
    u<-exp.mean(100)
    n<-c(n,u)
    x=x+1
    if(x>1000){break}
}
par(mfrow=c(1,2))
plot(density(m),xlim=c(0,15),ylim=c(0,1)) #size =10 , on the left side
plot(density(n),xlim=c(0,15),ylim=c(0,1)) #size =100 ,on the right side
```



We **choose exponential distribution** and notice that the curves are pretty similar to those we got in part c

4 Problem 3

Let X be a random variable for the face value of a fair d-sided die after a single roll. X follows a discrete uniform distribution of the form unif$\{1,d\}$. Below is the mean and variance of unif$\{1,d\}$.

$$E[X] = \frac{1+d}{2} \qquad Var(X) = \frac{(d-1+1)^2-1}{12}$$

a) Let $\bar{X}_n$ be the random variable for the mean of n die rolls. Based on the Central Limit Theorem, what distribution does $\bar{X}_n$ follow when $d = 6$.

$$X_n \sim \left(\mu, \frac{\sigma^2}{n}\right)$$

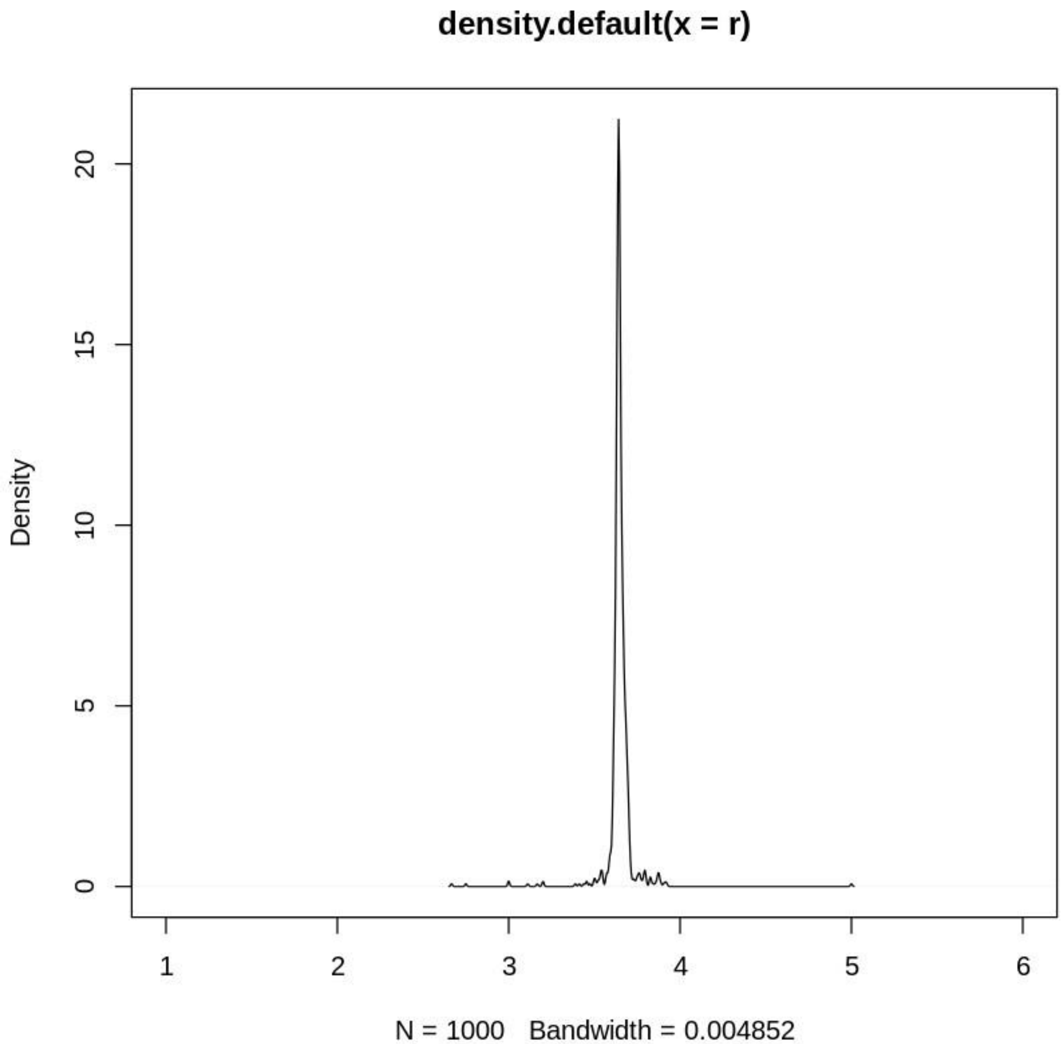
$$X_n \sim \left(3.5, \frac{(\frac{35}{12})}{n}\right)$$

b) Generate $n = 1000$ die values, with $d = 6$. Calculate the running average of your die rolls. In other words, create an array r such that:

$$r[j] = \sum_{i=1}^{j} \frac{X_i}{j}$$

Finally, plot your running average per the number of iterations. What do you notice?

[36]:
```r
f=vector()
r=vector()
for(i in 1:1000)# n=1000
{
s<-sample(1:6,1)
f=c(f,s)# f stores outcomes from each sim
r[i]=sum(f)/i
}
plot(density(r),xlim=c(1,6))
```



It looks like r follows a normal distribution