P_M6_1

September 4, 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.

$E[W] = E[X] + E[Y] = + = 2$

$Var(W) = Var(X) + Var(Y) = ^2 + ^2 = 2^2$

Then, by the central limit theorem,

$W\ N(2, 2^2)$

Also, we have

$E[U] = E[2X] = 2E[X] = 2$ and $Var(U) = Var(2X) = 2^2Var(X) = 4^2$

So by CLT

$U\ N(2, 4^2)$

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 $Var(X)$?

$E[X]$ = (a + b)/2 = (100 + 0)/2 = 50

$V[X] = (b - a)^2/12 = (100 - 0)^2/12 = 10000/12 = 833.33$

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?

[8]:
```r
uniform.sample.mean = function(n){

    samples = runif(n, 0, 100)
    sample.mean = mean(samples)

    return(sample.mean)
}

uniform.sample.mean(10)
uniform.sample.mean(100)
uniform.sample.mean(1000)
uniform.sample.mean(10000)
```

48.1728791072965

46.9700699464884

49.5594291877234

49.9760803943849

As we increase the size of n, then yes this becomes a good approximation of the population mean.

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

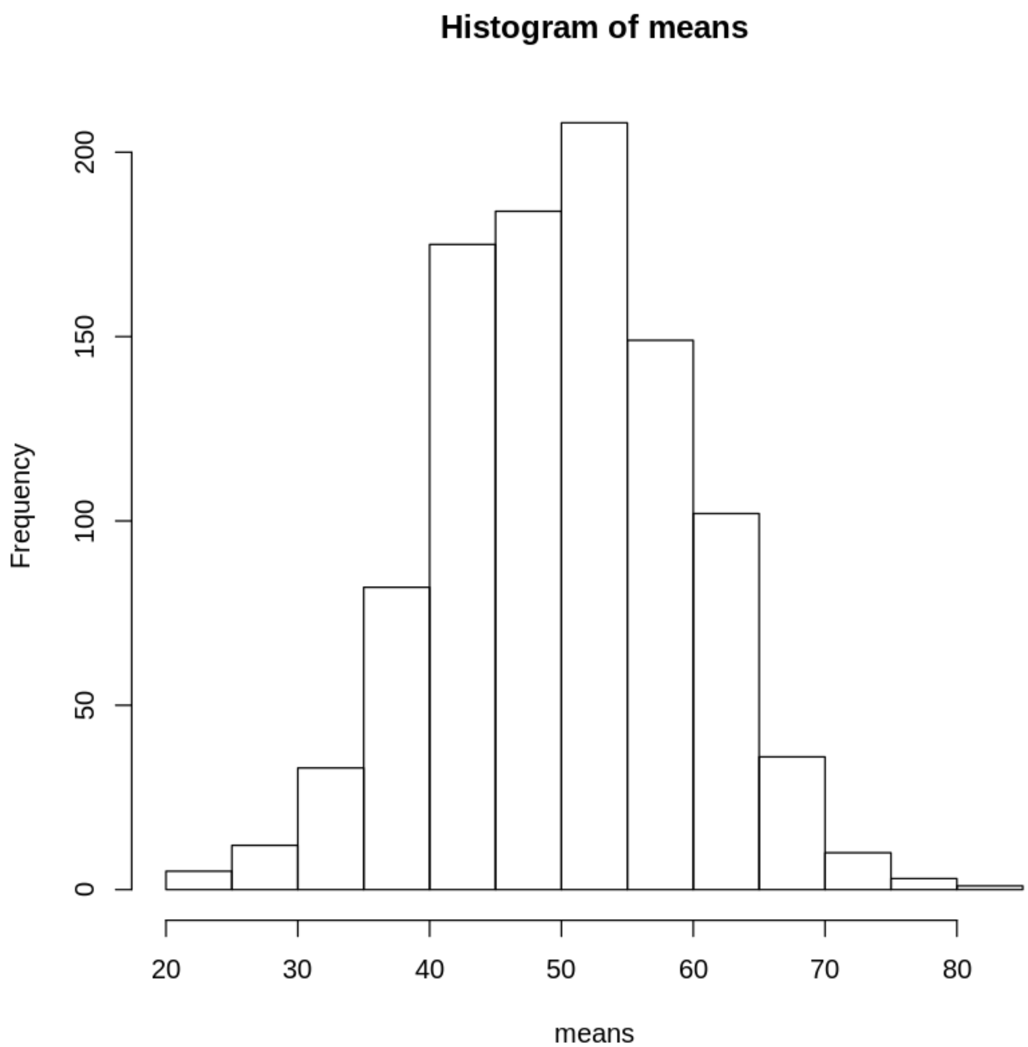
$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

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.

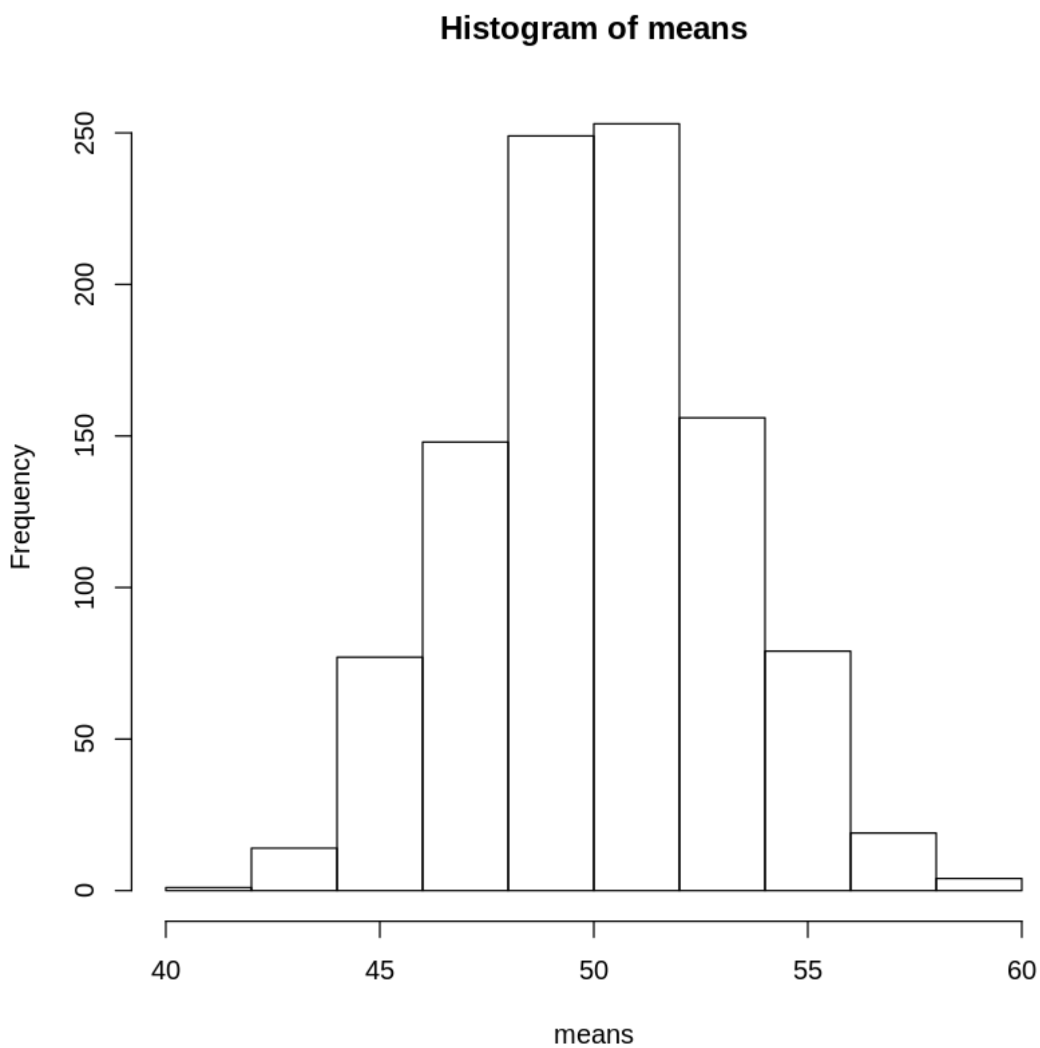
Using the `uniform.sample.mean` function, simulate $m = 1000$ sample means, each from a sample of size $n = 10$. Create a histogram of these sample means. Then increase the value of n and plot the histogram of those sample means. What do you notice about the distribution of $\bar{X}$? What is the mean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

[10]:
```r
means = numeric(1000)

for (i in 1:1000){
        means[i] = uniform.sample.mean(10)
}
```

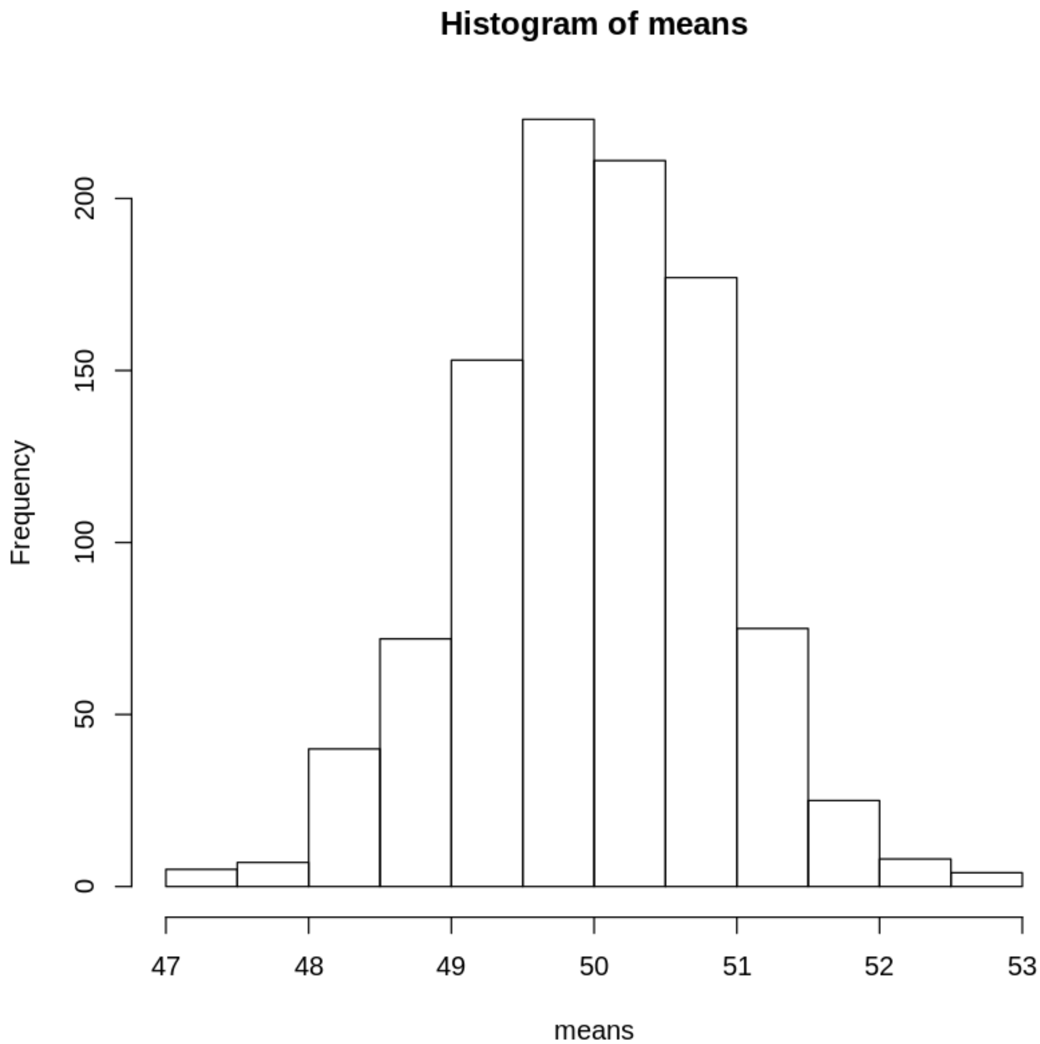
```
hist(means)
```




```
[11]: means = numeric(1000)

      for (i in 1:1000){
              means[i] = uniform.sample.mean(100)
      }

      hist(means)
```



[12]:
```
means = numeric(1000)

for (i in 1:1000){
        means[i] = uniform.sample.mean(1000)
}

hist(means)
```

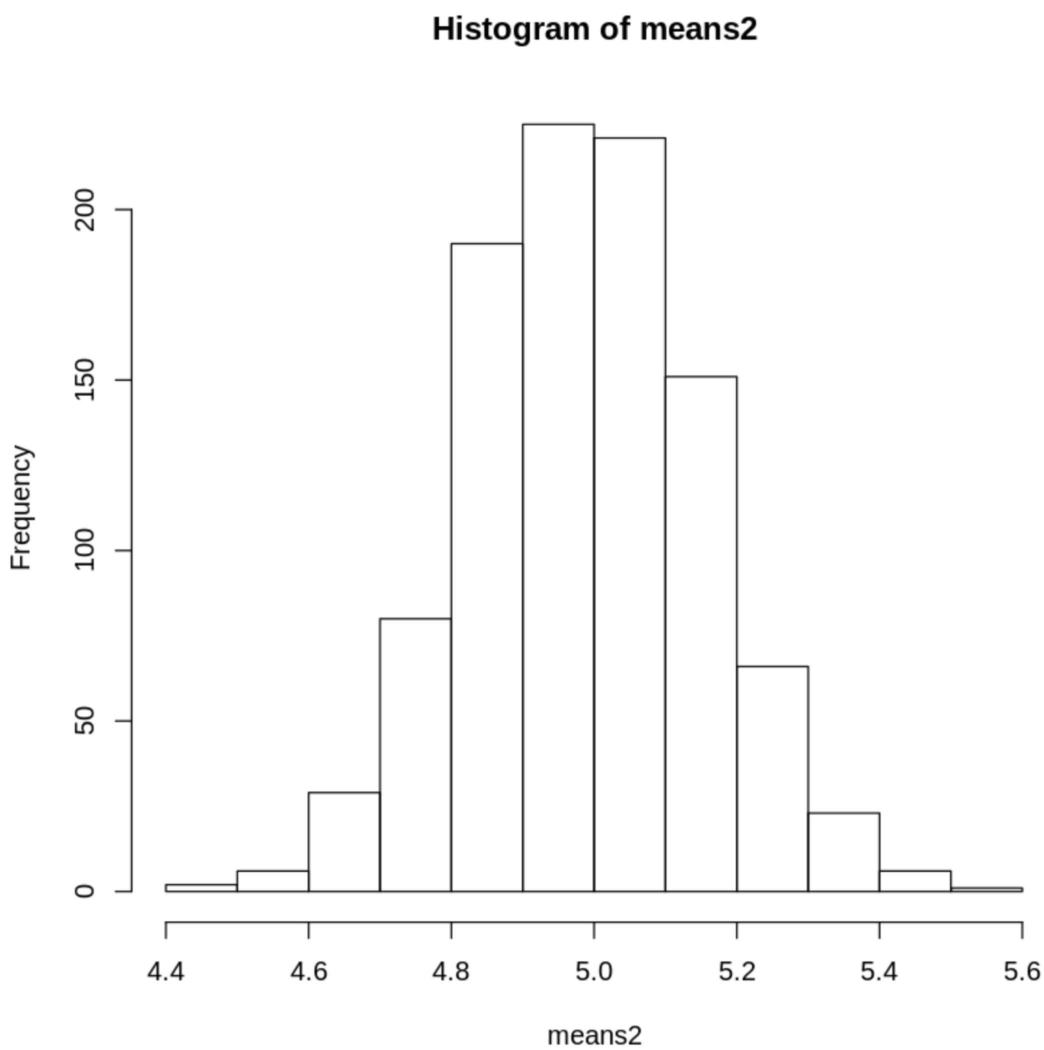


Distribution of x-bar is normal, the distribution is always centered around 50, and as n increased, the variance of x-bar decreases.

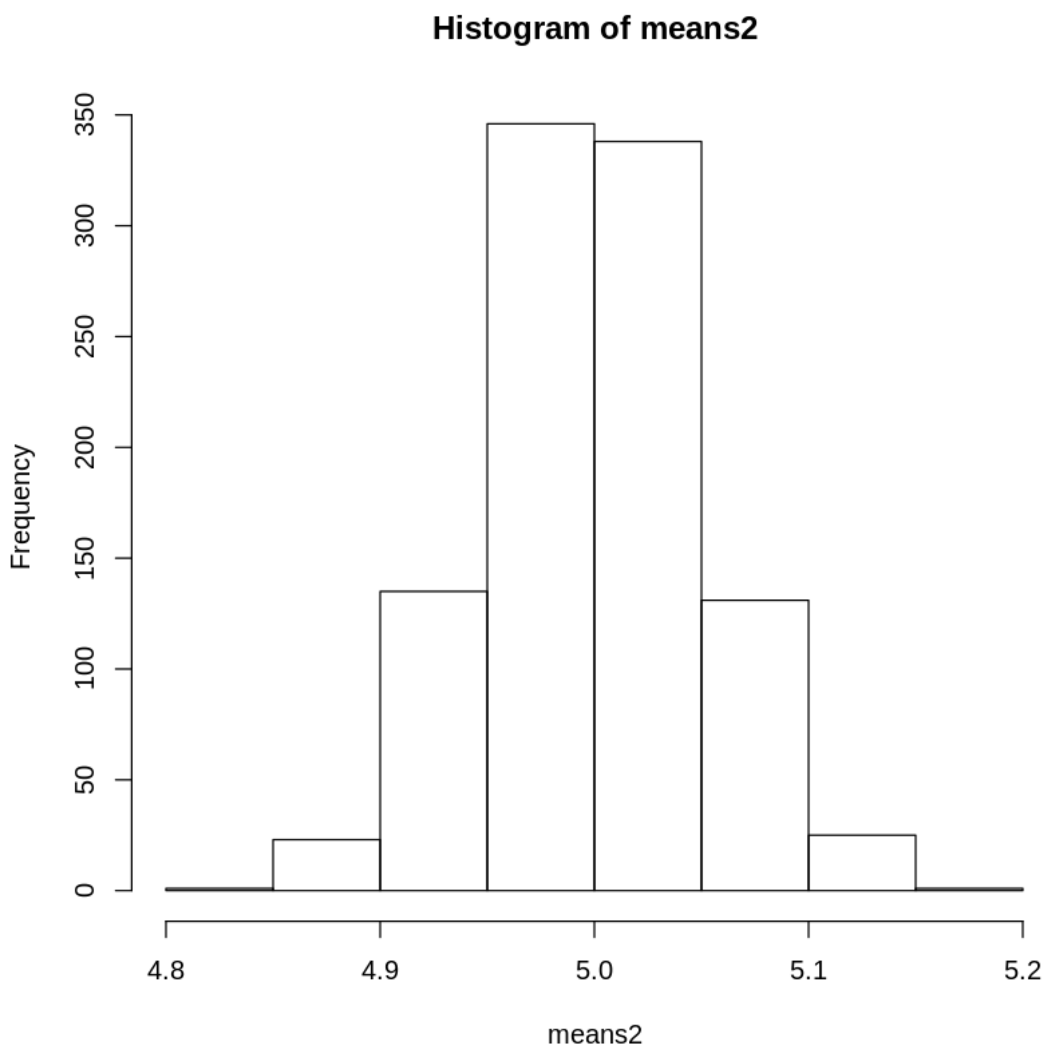
d) Recall that our underlying population distribution is $U(0, 100)$. Try changing the underlying distribution (For example a binomial(10, 0.5)) and check the sample distribution. Be sure to explain what you notice.

[19]:
```
binomial.sample.mean = function(n){
    total = 0
    for(i in 1:n){
        dist = rbinom(1, 10, 0.5)
        total = total + dist
    }
```

```
    sample.mean = total / n

    return (sample.mean)
}

means2 = numeric(1000)
for (i in 1:1000){
    means2[i] = binomial.sample.mean(100)
}

hist(means2)
```



[20]:
```
means2 = numeric(1000)
for (i in 1:1000){
    means2[i] = binomial.sample.mean(1000)
}

hist(means2)
```



Sample distribution is normal, as n increases variance decreases, and the sample mean is centered at the theoretical mean.

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 $d = 6$.

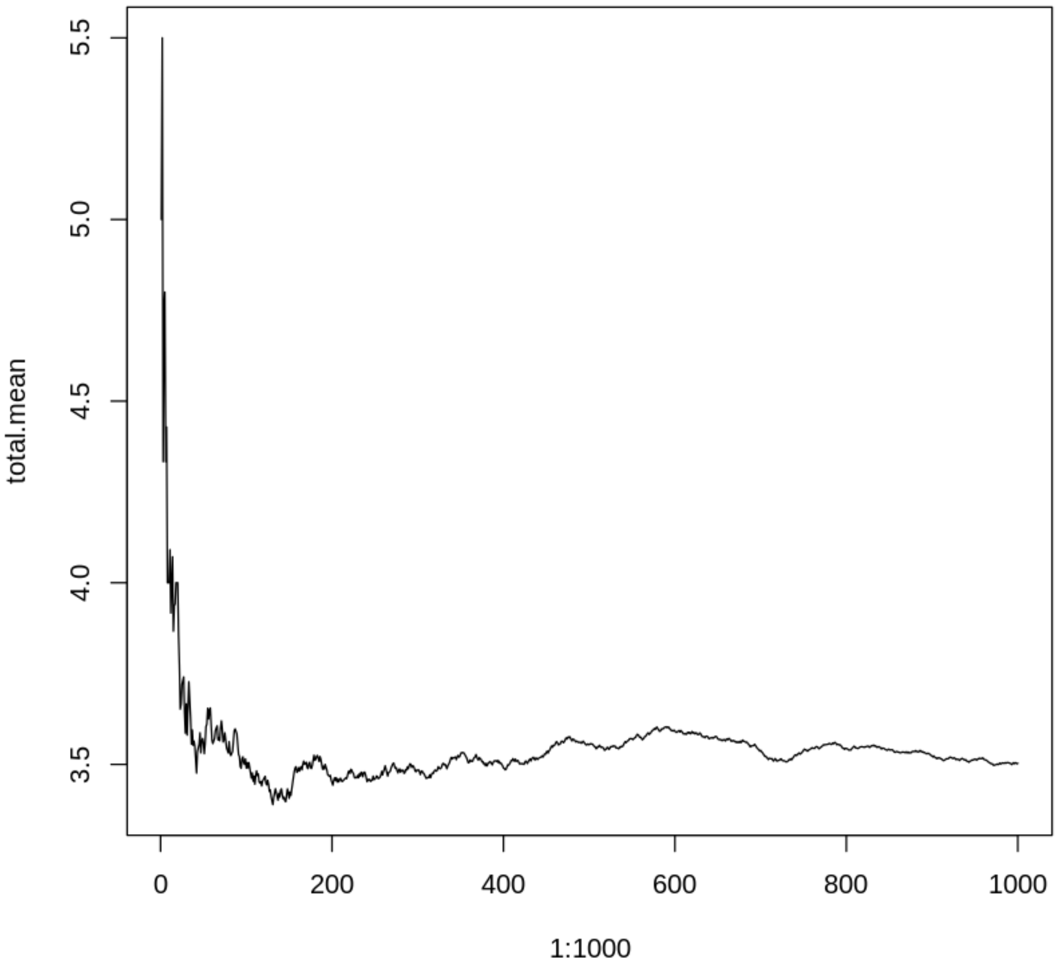
x-bar ~ N((1+6)/2, (6-1+1)^2 / 12) = N(3.5, 2.92/n)

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?

[21]:
```
rolls = sample(1:6, size=1000, replace=TRUE)
total.rolls = cumsum(rolls)
total.mean = total.rolls / (1:1000)
plot(x=1:1000, y=total.mean, type="l")
```



As the number of iterations increases, the calculated average approaches the theoretical mean.

[ ]: