Assignment 1, Random Walk

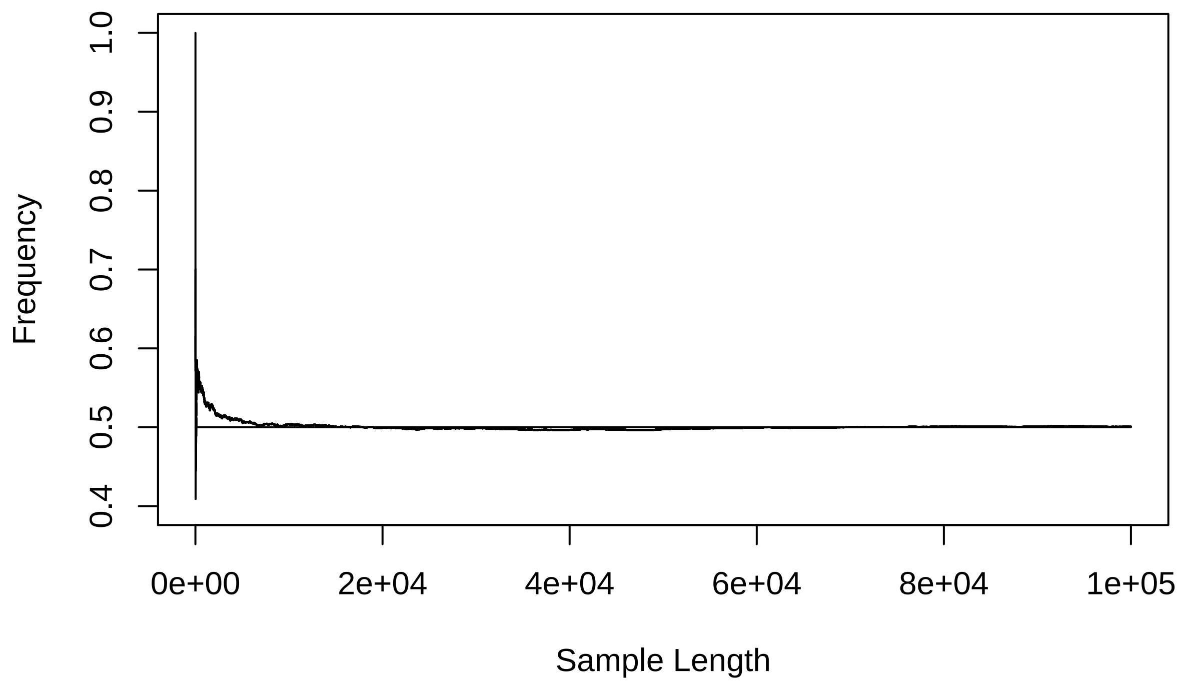
Matthew Dunne

April 1, 2018

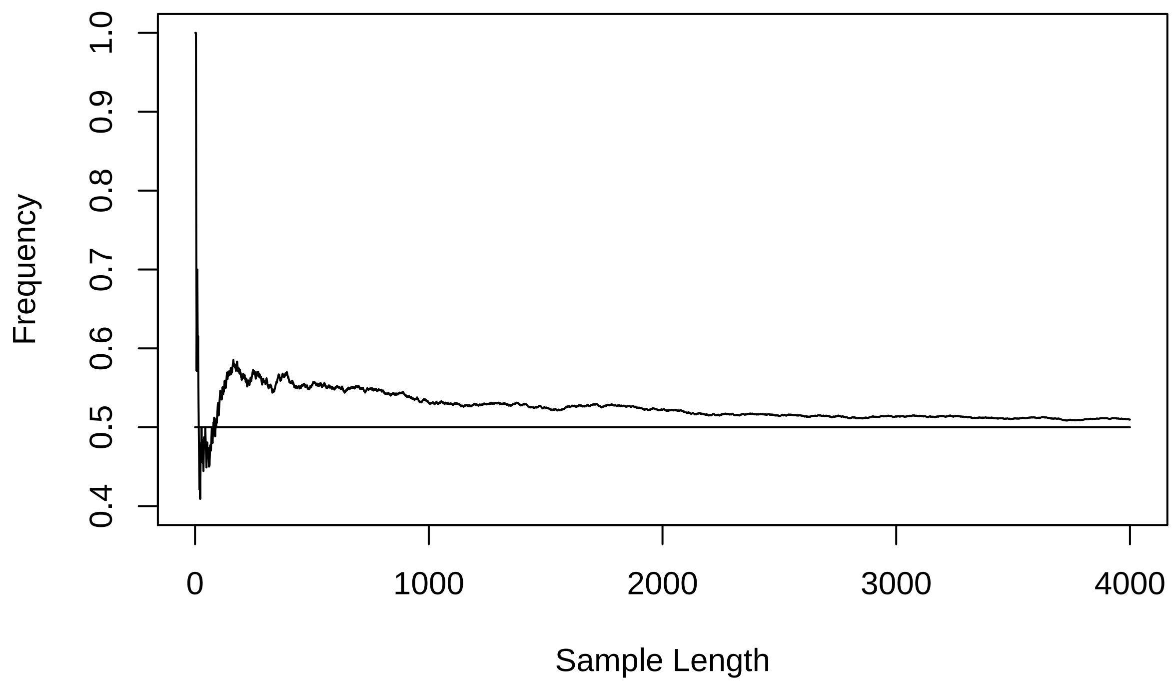
1. Convergence of probability of tail to 0.5

Check that frequency of "Tails" (outcome equals 1) converges to 0.5 as the number of tosses grows. Simulate a sample of 100,000 flips of a presumably fair coin.

```
#set seed to ensure reproducibility
set.seed(12345)
#set number of flips to 100,000
nFlips<-100000
#create a vector that randomly chooses between 0 and 1 100,000 times, with replacement
Flips<-sample(0:1,nFlips,repl=T)
# or rbinom(nFlips,1,.5) or (runif(nFlips)<.5)
#get the cumulative sum of those 100,000 flips, i.e. add up all the 1's
Trajectory<-cumsum(Flips)
#find out what percentage of the vector are 1's
freq<-Trajectory/(1:nFlips)
#plot a graph, x-axis is the length of the orginal vector (# of flips), y-axis is the cumulative percen
plot(1:length(freq),freq, ylim=c(.4,1),type="l",ylab="Frequency",xlab="Sample Length")
#draw a line at 50% on the y-axis
lines(c(0,nFlips),c(.5,.5))
```




```
#plot the same graph as above except focus on the first 4000 flips
plot(1:4000,freq[1:4000], ylim=c(.4,1),type="l",ylab="Frequency",xlab="Sample Length")
lines(c(0,4000),c(.5,.5))
```



What does this say about fairness of the coin?
It appears that as the number of flips increases, the frequency of tails goes to 0.50. It swings wildly in the early flips, probably between 1 and 0 (we only see 0.4 to 1 in these graphs). In the second graph it shows that the coin remains ever so slightly in favor of tails (above the line) for a significant number of flips.

2. Check your intuition about Random Walks

2.1 - One Trajectory

Create a trajectory of wealth in a game which either pays $1 with probability 0.5 or results in loss of $1 on each step. Assume that the game is played 1,000,000 times.

```
#set number of flips to 1 million
nFlips<-1000000;
#set seed for reproducibility
set.seed(12345)
#generate a vector of flips, convert from 0 or 1 to a payout of -1 or 1.
Flips<-(sample(0:1,nFlips,repl=T)-.5)*2
```

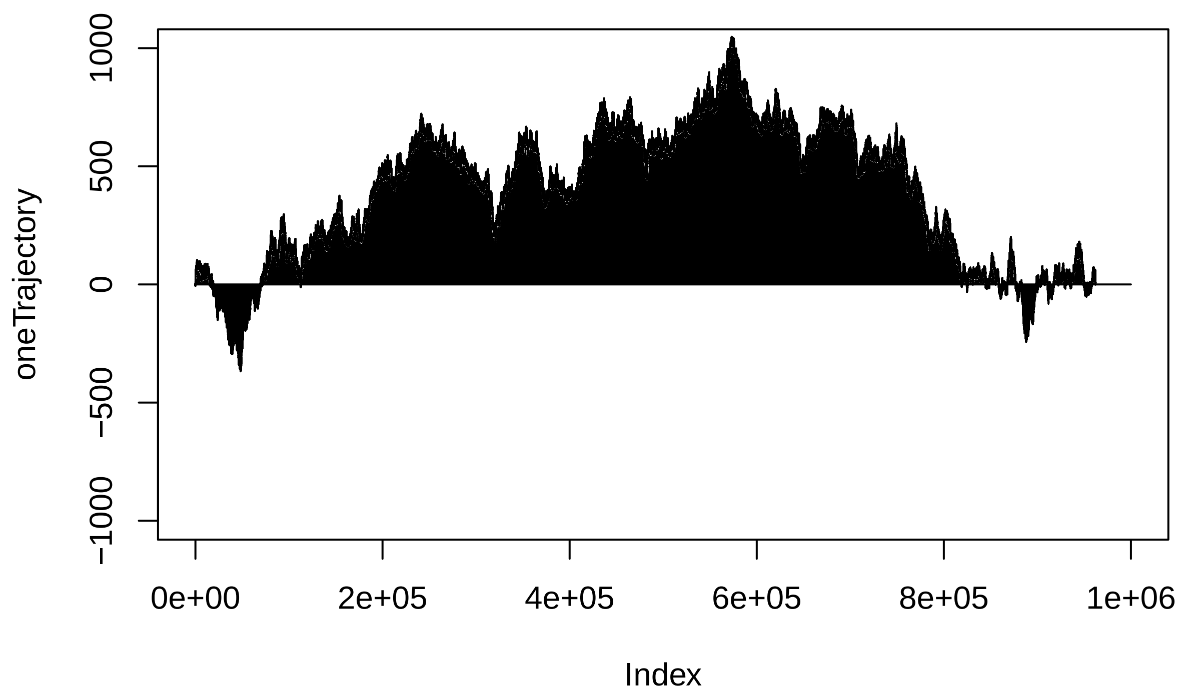
An alternative way of creating this trajectory would be:

```
#set number of flips to 1 million
nFlips<-1000000;
#set seed for reproducibility
set.seed(12345)
#generate a vector of flips, convert from 0 or 1 to a payout of -1 or 1.
AltFlips<-(rbinom(nFlips,1,.5)-.5)*2
```

Before plotting the trajectory of wealth, it is useful to guess what that trajectory will look like. In the opinion of this author, we will likely see a line that rises and falls above $0, being above it and below it in equal measure, and not straying very far from $0.

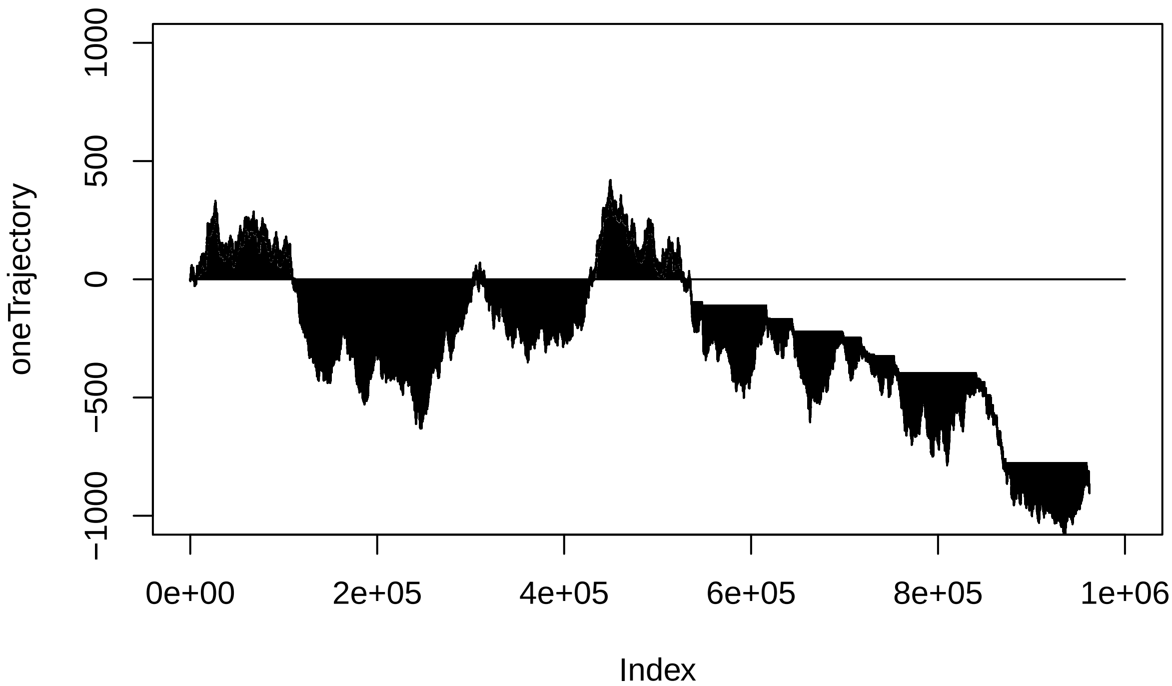
Here is the plot of the trajectory of wealth:

```
#get the cumulative sum from all the -1 and 1 flips
oneTrajectory<-cumsum(Flips)
#plot the cumulative sum
plot(oneTrajectory, ylim=c(-1000,1000),type="l")
#add a horizontal line at $0 in wealth for refrence
lines(c(0,nFlips),c(0,0))
```

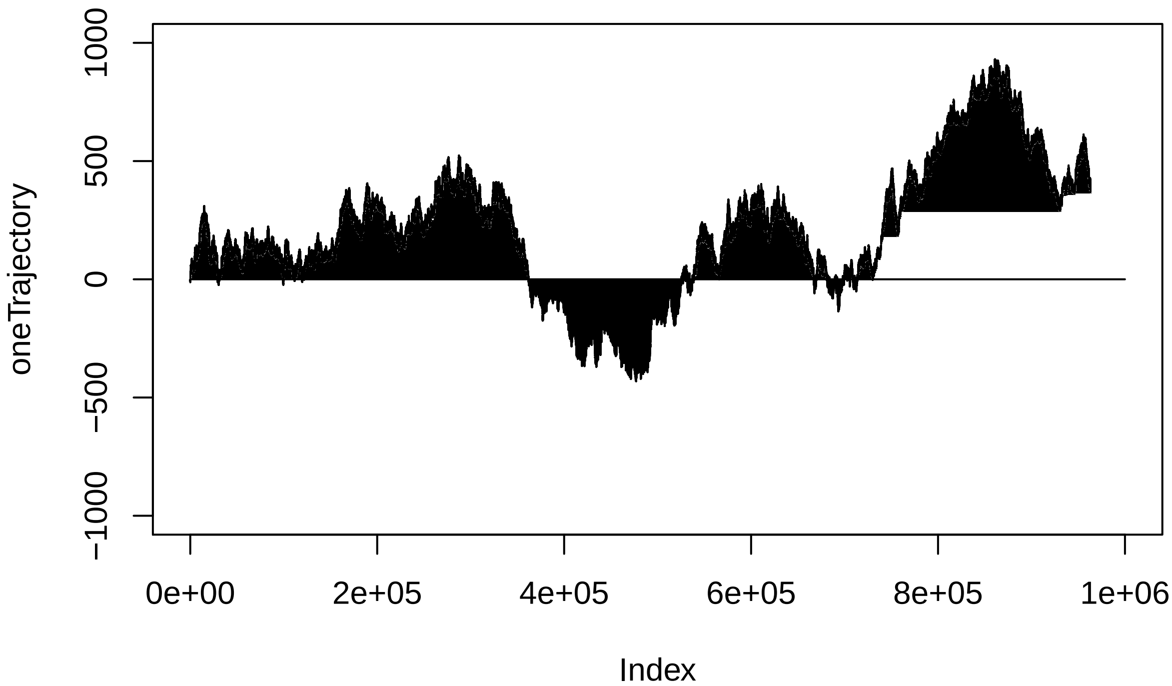


We can repeat this experiment a few more times.

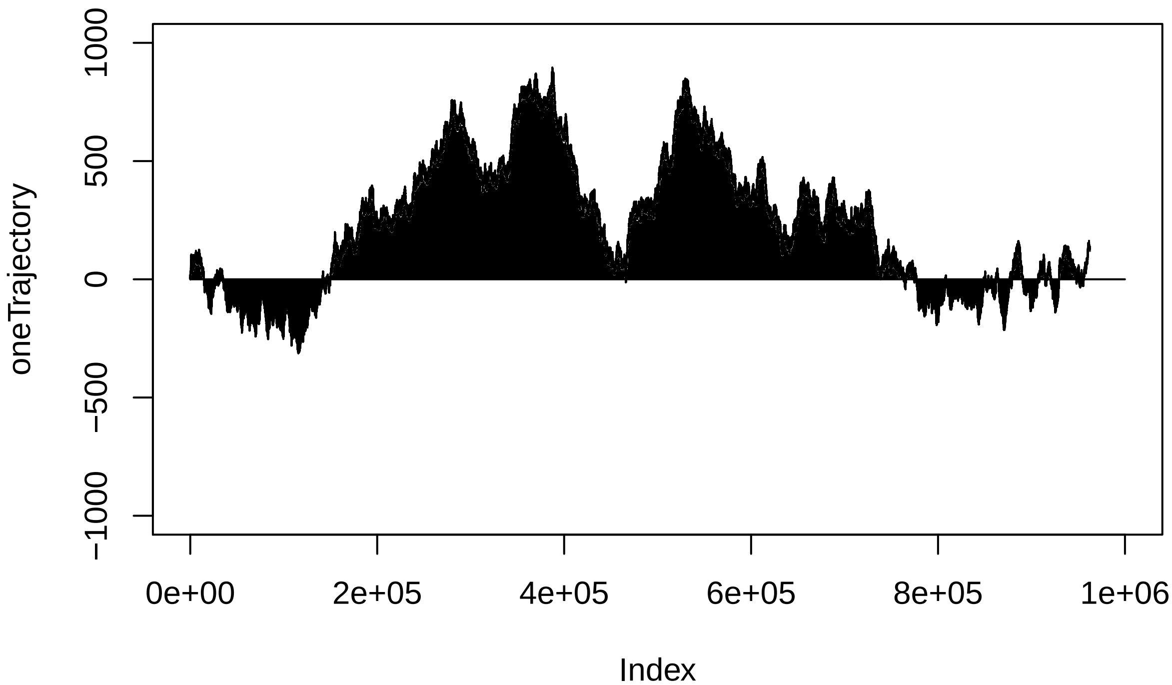
```
nFlips<-1000000;
#set seed for reproducibility
#set.seed(12345)
#generate a vector of flips, convert from 0 or 1 to a payout of -1 or 1.
AltFlips<-(rbinom(nFlips,1,.5)-.5)*2
#get the cumulative sum from all the -1 and 1 flips
oneTrajectory<-cumsum(AltFlips)
#plot the cumulative sum
plot(oneTrajectory, ylim=c(-1000,1000),type="l")
#add a horizontal line at $0 in wealth for refrence
lines(c(0,nFlips),c(0,0))
```




```r
nFlips<-1000000;
#set seed for reproducibility
#set.seed(12345)
#generate a vector of flips, convert from 0 or 1 to a payout of -1 or 1.
AltFlips<-(rbinom(nFlips,1,.5)-.5)*2
#get the cumulative sum from all the -1 and 1 flips
oneTrajectory<-cumsum(AltFlips)
#plot the cumulative sum
plot(oneTrajectory, ylim=c(-1000,1000),type="l")
#add a horizontal line at $0 in wealth for refrence
lines(c(0,nFlips),c(0,0))
```




```r
nFlips<-1000000;
#set seed for reproducibility
#set.seed(12345)
#generate a vector of flips, convert from 0 or 1 to a payout of -1 or 1.
AltFlips<-(rbinom(nFlips,1,.5)-.5)*2
#get the cumulative sum from all the -1 and 1 flips
oneTrajectory<-cumsum(AltFlips)
#plot the cumulative sum
plot(oneTrajectory, ylim=c(-1000,1000),type="l")
#add a horizontal line at $0 in wealth for refrence
lines(c(0,nFlips),c(0,0))
```



Do these simulation match my expectations? They do not. Each simulation is different, but in some we see significantly more time spent above or below $0 than I would have expected. Also the payouts, both negative and positive, are larger than I expected, reaching almost $1000 in each direction.

2.2 - Multiple Trajectories

What do you expect the following probabilities to be?
-In 500 coin flips, the probability that the difference in number of heads and number of tails is less than 5
-In 500 coin flips, the probability that the difference in number of heads and number of tails is greater than 25

I would expect the probability that the difference is less than 5 is about 2% and would expect that the difference is greater than 25 is about 20%.

To test this, we will turn the 1,000,000 coin flips into 2000 random walk samples, each is 500 long and then calculate 2000 cumulative trajectories. Each trajectory at each point shows the difference between the number of "Tails" and the number of "Heads".

We will find how many times out of 2,000 runs:
-Trajectories end less than 5 points away from zero (5 is 1% of 500 tosses)
-Trajectories end more than 25 points away from zero (25 is 5% of 500 tosses)
and then estimate the probabilities of such deviations.

```r
#make the vector of 1 million flips into a 2000 by 500 matrix, 2000 runs of 500 flips keeping track of
Trajectories2000by500<-t(apply(matrix(Flips,ncol=500),1,cumsum))
#check to make sure the dimensions are 2000 by 500
dim(Trajectories2000by500)
```

```
## [1] 2000  500
```

```r
#calculate the percentage of rows (trials) where the cumulative sum is less than 5 away from 0
(probability.less.than.5<-sum(abs(Trajectories2000by500[,500])<5)/2000)
```

```
## [1] 0.18
```

```r
#calculate the percentage of rows (trials) where the cumulative sum is more than 25 away from 0
(probability.greater.than.25<-sum(abs(Trajectories2000by500[,500])>=25)/2000)
```

```
## [1] 0.2515
```

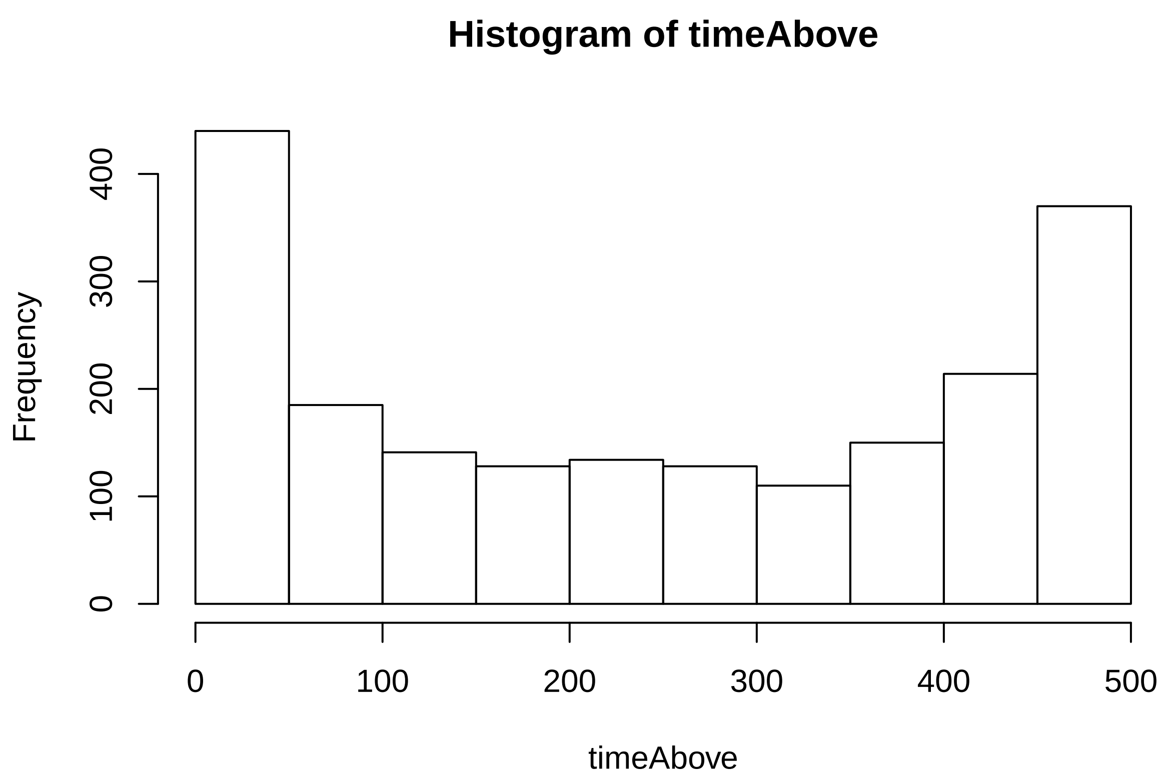
We see that the probability that the trajectory will end less than 5 points away from 0 is 18%, and the probability that it will end more than 25 points away is 25.15%. I was far off in my estimate of the former and somewhat close in my estimate of the latter.

2.3 - Time on One Side

How long do you expect trajectory of a random walk to spend on one side from zero, below or above? I would expect the trajectory of a random walk to spend as much time above 0 as below, a 50/50 split.

We can look at each of the 2000 trials of 500 flips and calculate, for each, how many of those 500 flips were done when the cumulative value was above 0:

```r
#for each of the 2000 tests find in how many of their 500 flips the cumulative value was above 0
timeAbove<-apply(Trajectories2000by500,1,function(z) sum(z>0))
#plot a histogram of that calculation.
hist(timeAbove)
```



What sticks out is the large number of trials at the extremes. There are about 450 trials where between 0 and 50 flips were spent above 0 and about 400 trials where between 450 and 500 flips were spent above 0.

Furthermore, there is some symmetry in between those extremes, with the bars of the histogram decreasing precipitously, then gradually, then increasing gradually, and then precipitously.

As such, I can say my intuition was not correct. If random walks spent as much time above 0 as below one would expect the histogram to look something like a normal distribution with a mean of 250. Instead we see almost the opposite.

The law that we observe in this histogram is the arcsine distribution law, or more precisely the First Arcsine law, which states that the proportion of time that the one-dimensional random walk is positive follows an arcsine distribution (probability distribution).