

Probability exercises

SMWadgymar

2022-08-16

####Exercises

- 1) Simulating coin flips: Generate 10 separate random flips with probability 0.7 of producing heads.

```
rbinom(10, 1, .7)
```

```
## [1] 0 1 1 1 0 1 1 1 0 0
```

Q: What kind of values do you see and what do they represent?

A:

- 2) Simulating draws from a binomial distribution: Generate 100 occurrences of flipping 10 coins, each with 70% probability of producing heads. Produce a table of the probability distribution that you generate (hint: your table should report probabilities and not counts).

```
flips<-rbinom(100, 10, 0.7)
```

```
prop.table(table(flips))
```

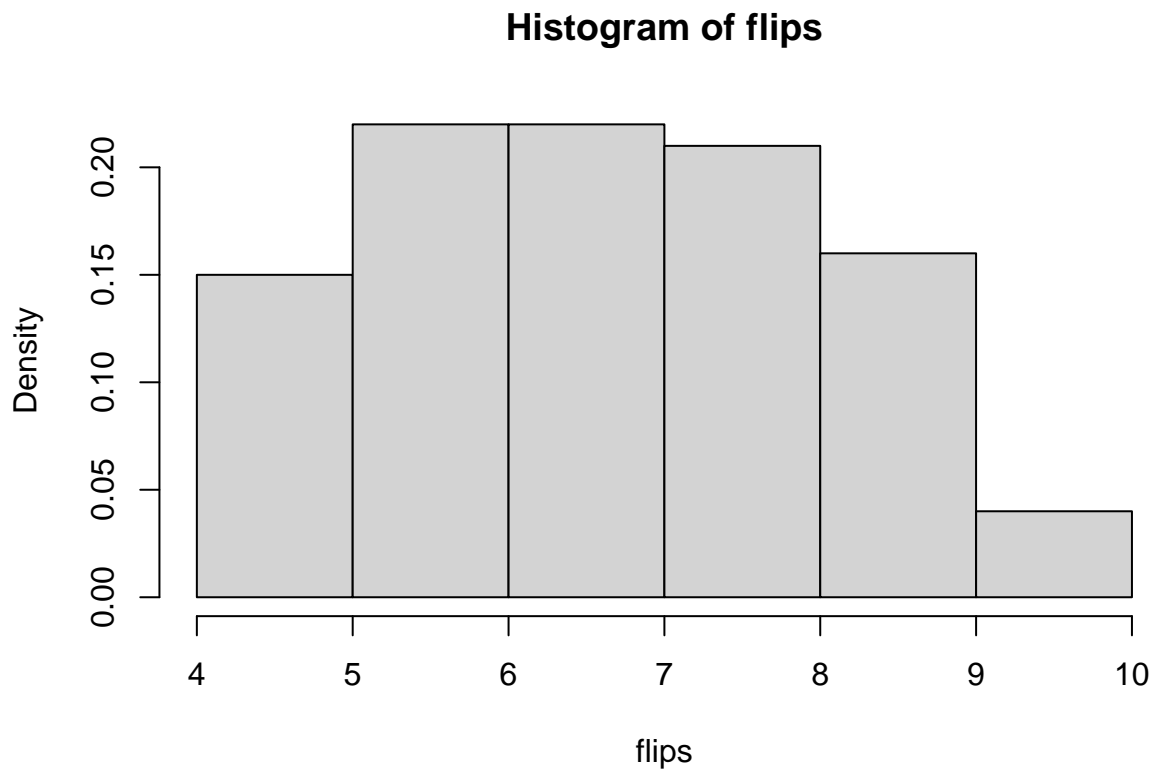
```
## flips
##    4    5    6    7    8    9   10
## 0.02 0.13 0.22 0.22 0.21 0.16 0.04
```

```
#Could also use table(flips)/100 to get the table
```

Q: What kind of values do you produce and what do they represent? Plot a histogram and describe the shape that your probability distribution takes. What does the freq=F argument mean?

A:

```
hist(flips, freq=F)
```



- 3) Calculate the exact probability that 2 heads will arise from 10 coin flips with a 70% probability of coming up tails.

```
dbinom(2, 10, 0.3)
```

```
## [1] 0.2334744
```

#Note that the word problem lists the probability of getting tails

Q: Compare your answer with a simulation of 10,000 trials. Do the two approaches yield similar results?

A:

```
mean(rbinom(10000, 10, 0.3)==2)
```

```
## [1] 0.2282
```

- 4) Calculate the cumulative probability that at least five coins out of 10 are heads with a 30% probability of coming up heads.

```
1-pbinom(4, 10, 0.3)
```

```
## [1] 0.1502683
```

##Note that we have to subtract from 1 to get the right range

Q: Compare your answer with a simulation of 10,000 trials. Do the two approaches yield similar results?

A:

```
mean(rbinom(10000, 10, 0.3)>= 5)
```

```
## [1] 0.1447
```

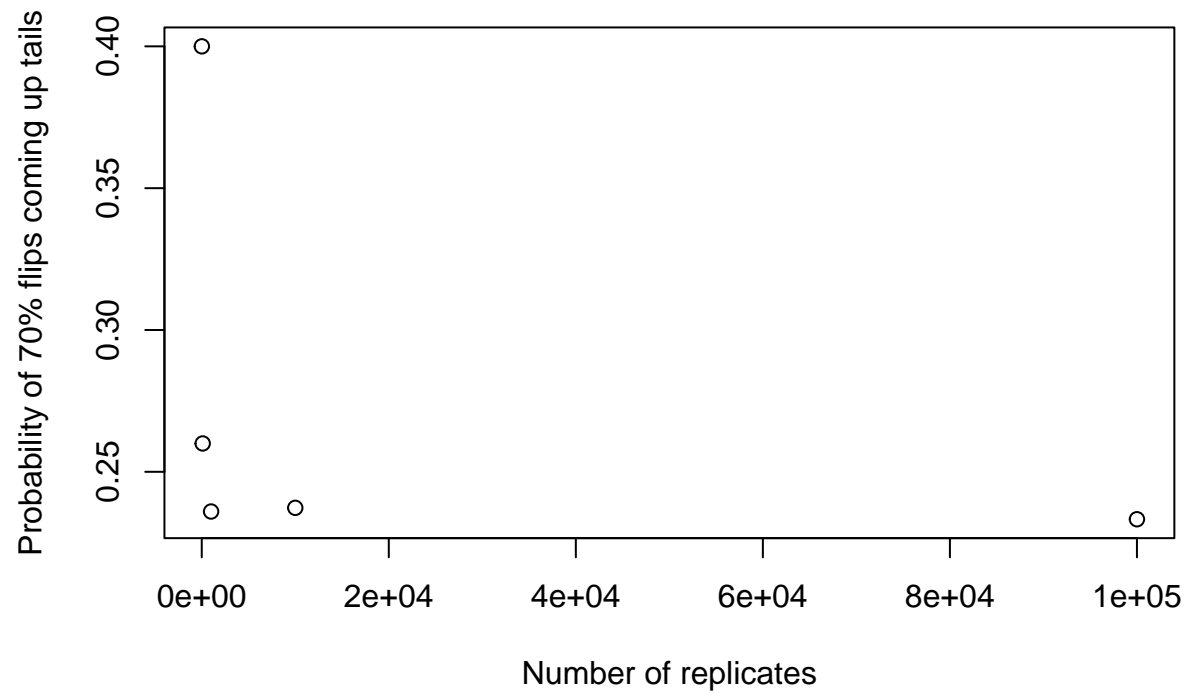
- 5) Repeat the simulation you ran in exercise (3) where you estimated the exact probability that two heads will occur with 10, 100, 1,000, 10,000, and 100,000 trials. Produce a plot depicting the number of trials on the x-axis and the associated probabilities you calculated on the y-axis.

Hint: The `c()` function lets you combine values into a vector or list. You will need to make a vector of probabilities for the y axis using `and` a vector of your sample sizes for the x axis using `c(#, #, #, #, #)`.

```
r10<-mean(rbinom(10, 10, 0.3)==2)
r100<-mean(rbinom(100, 10, 0.3)==2)
r1000<-mean(rbinom(1000, 10, 0.3)==2)
r10000<-mean(rbinom(10000, 10, 0.3)==2)
r100000<-mean(rbinom(100000, 10, 0.3)==2)

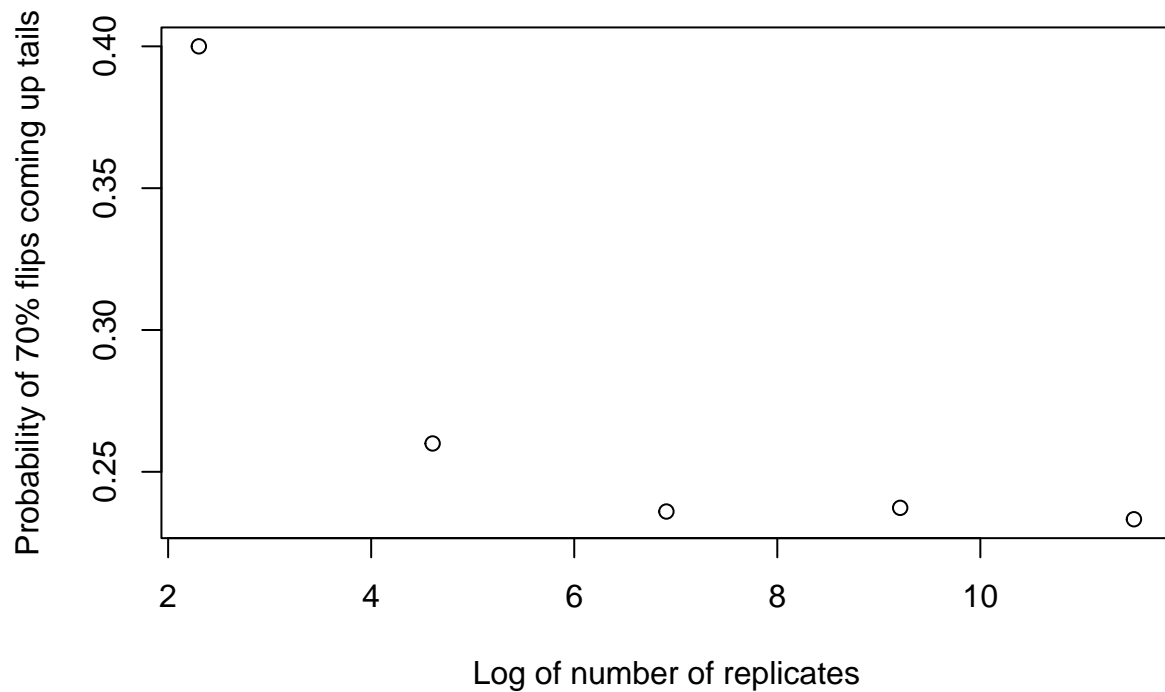
r<-c(r10, r100, r1000, r10000, r100000)
n<-c(10, 100, 1000, 10000, 100000)

plot(n, r, xlab="Number of replicates", ylab="Probability of 70% flips coming up tails")
```



Produce a second plot but this time log-transform the number of trials. Make sure to adjust your axis labels appropriately.

```
plot(log(n), r, xlab="Log of number of replicates", ylab="Probability of 70% flips coming up tails")
```



Q: Which simulation yields a result most similar to the exact probability? What pattern do you see in your first graph? What does your second graph reveal that your first graph did not?

A: The simulation with the largest sample size yields a result most similar to the exact probability. The sample sizes differ by orders of magnitude, which makes a log scale more appropriate for graphing. Once you log-transform the number of replicates, you can see that sample sizes of 100 and larger start to converge on the same value (the exact probability, which you calculated in question 3 using `dbinom(2, 10, 0.3)`).