IST772 Summary Template: Chapter 2 – Reasoning with Probability

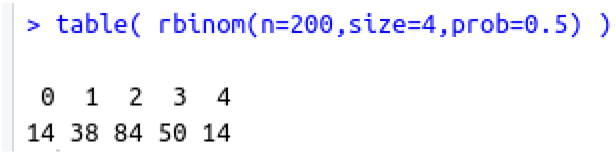
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Hendi Kushta
Date: 01/22/2023

****Important** Copying and/or pasting anything from the textbook will not be acceptable for your chapter notes submissions. You must write your notes in your own words and generate your own code, results, and graphs in R. This is what forces your brain to process the material that you read.**

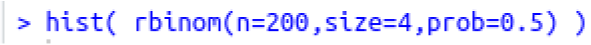
OUTCOME TABLES

An outcome table is the most basic and user-friendly tool for thinking about probability. It is a list of potential outcomes for a collection of related or comparable events.
Binomial distributions are what statisticians refer to when an event has two alternative outcomes, example the coin toss will be head or tail.
When the process of predicting an event is repeated many times, it is called a trial. While repeating the process many times, some events are predicted to occur more than the other.
In R to produce some random binomial distributions we use:



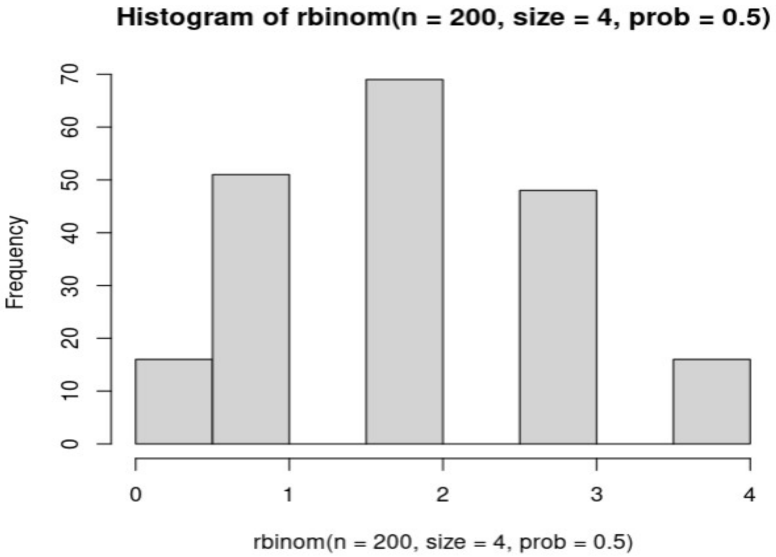
```
> table( rbinom(n=200,size=4,prob=0.5) )

 0  1  2  3  4 
14 38 84 50 14 
```

rbinom() function is used to generate some random binomial distributions, n = 200 tells to run the function 200 times, size = 4 tells function to be 4 predictions for each event and prob=0.5 tells that there is a 50% for each event to occur.
Table() function presents the data in a table.
Out of 200 trials, 84 are counted in the second trial, which is also the highest percentage of 84/200 = 0.42 or 42%. The histogram below also show the binomial distribution.



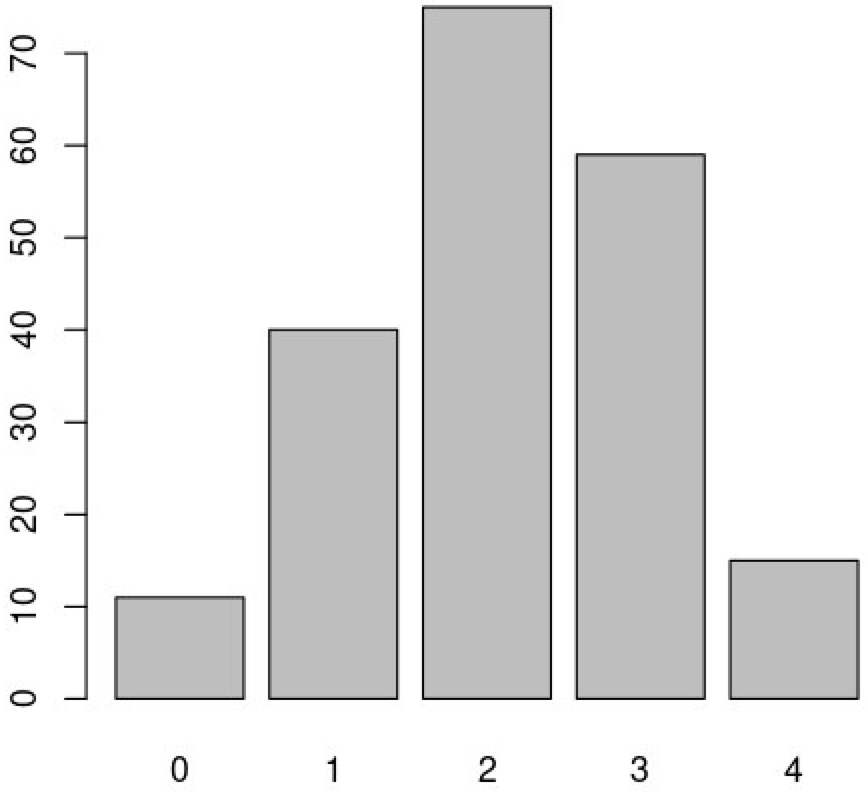
```
> hist( rbinom(n=200,size=4,prob=0.5) )
```

IST772 Summary Template: Chapter 2 – Reasoning with Probability
Originality Assertion: By submitting this file you affirm that this writing is your own.



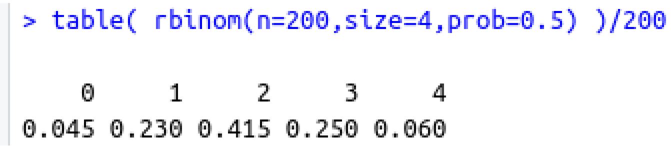
To get a better representation of the histogram above, we can use also barplot.

```
> barplot(table( rbinom(n=200,size=4,prob=0.5) ))
```

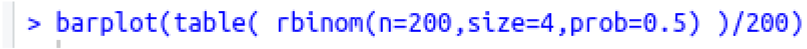


IST772 Summary Template: Chapter 2 – Reasoning with Probability
Originality Assertion: By submitting this file you affirm that this writing is your own.

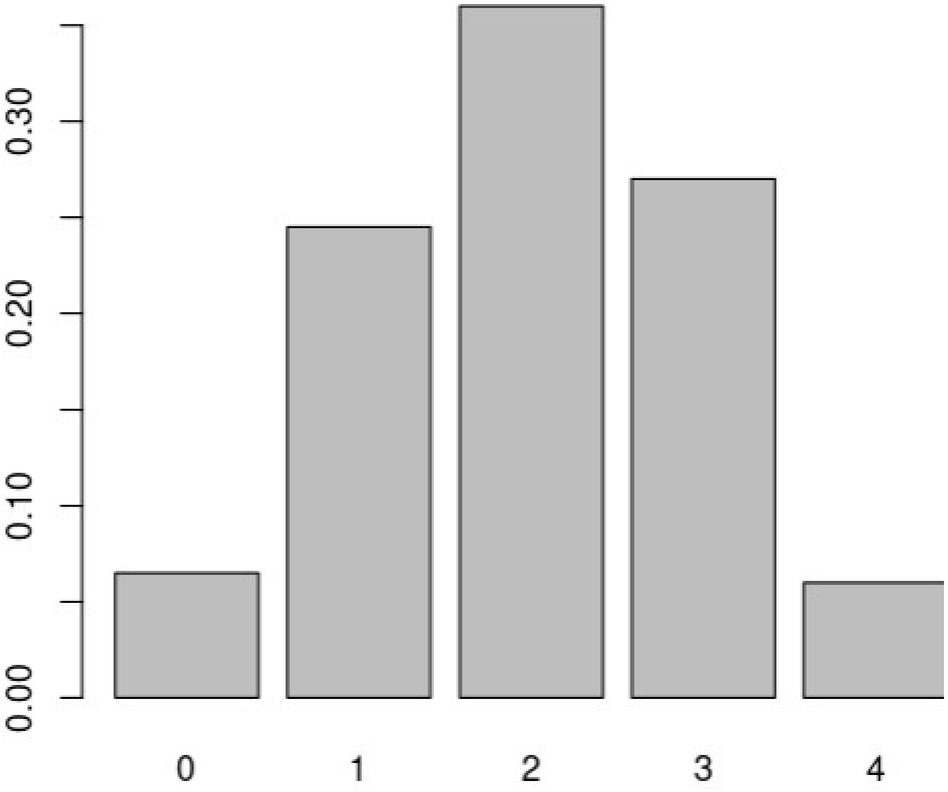
The same as I plotted the count of each occurrence, by using barplot, we can plot also the probability for each event by dividing each trial with the number total trials.



```
> table( rbinom(n=200,size=4,prob=0.5) )/200

    0     1     2     3     4 
0.045 0.230 0.415 0.250 0.060 
```



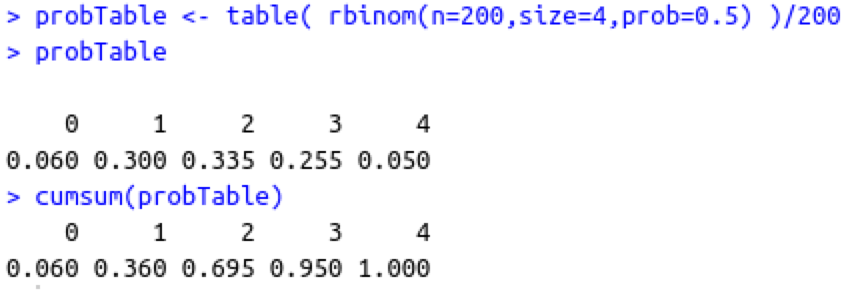
```
> barplot(table( rbinom(n=200,size=4,prob=0.5) )/200)
```



We can also find the cumulative probability when we want to combine events with one another, and build barplot for cumulative probability.

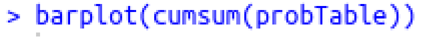
IST772 Summary Template: Chapter 2 – Reasoning with Probability

Originality Assertion: By submitting this file you affirm that this writing is your own.

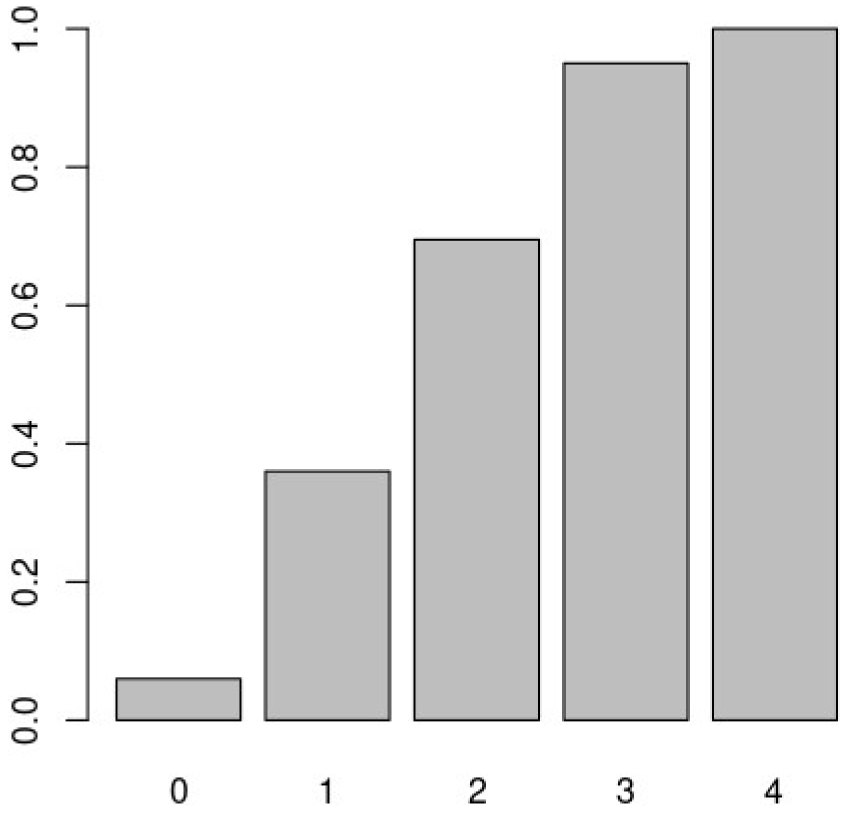


```
> probTable <- table( rbinom(n=200,size=4,prob=0.5) )/200
> probTable

    0     1     2     3     4 
0.060 0.300 0.335 0.255 0.050 
> cumsum(probTable)
    0     1     2     3     4 
0.060 0.360 0.695 0.950 1.000 
```

As shown, I have firstly found the probability for each event, then assigned to a variable named probtable, and in the end found the cumulative probability which should always add to 1. Below is the barplot for the cumulative probability.



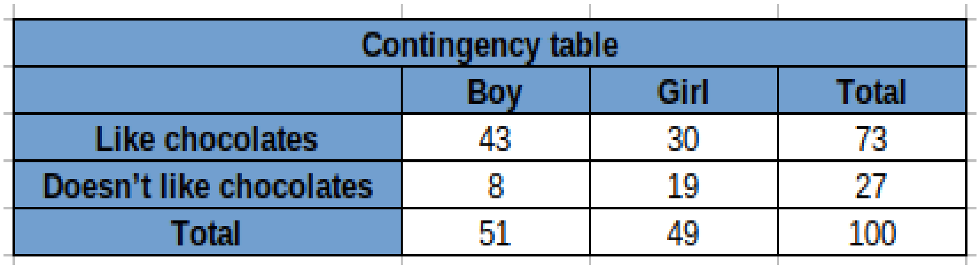
```
> barplot(cumsum(probTable))
```



IST772 Summary Template: Chapter 2 – Reasoning with Probability
Originality Assertion: By submitting this file you affirm that this writing is your own.

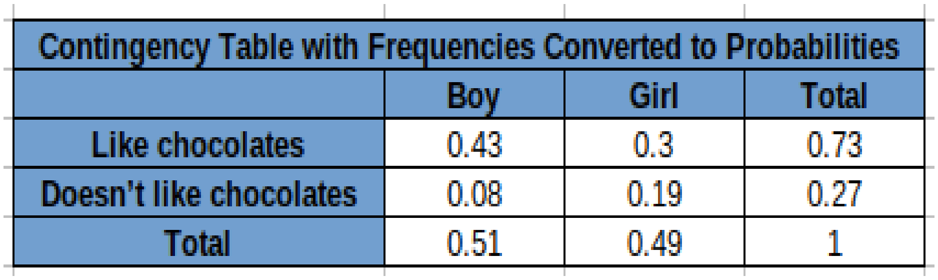
CONTINGENCY TABLES

A matrix-style table known as a contingency table shows the variables' (multivariate) frequency distribution. To be called a contingency table, it should have at least 4 cells.
Sometimes it is easier to add totals on the right and on the left of each column and row named also as marginal totals, and the grand total which is the number of total events.



Contingency table			
	Boy	Girl	Total
Like chocolates	43	30	73
Doesn't like chocolates	8	19	27
Total	51	49	100

Sum for each columns are 51 and 49 and sum of each rows are 73 and 27. The total sum is 100. Below I have shown the contingency table with the frequencies converted to probabilities.



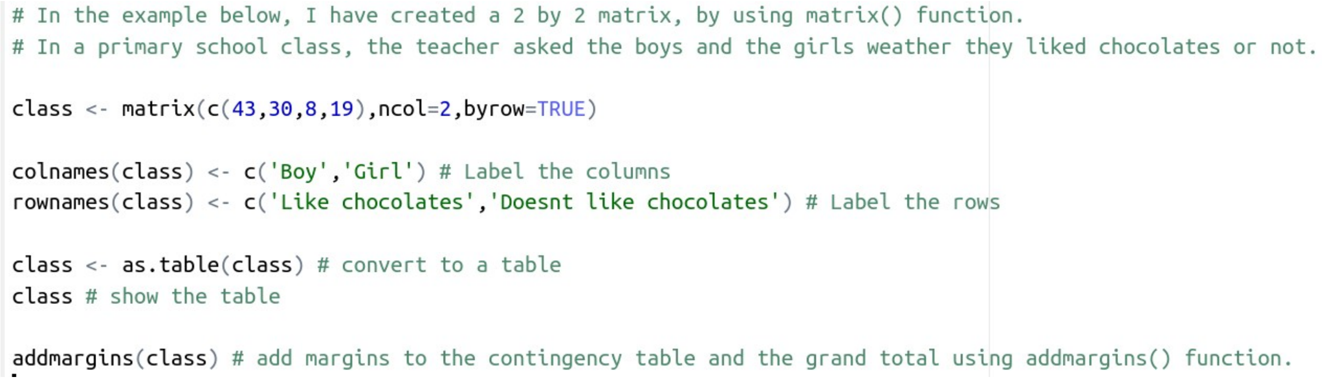
Contingency Table with Frequencies Converted to Probabilities			
	Boy	Girl	Total
Like chocolates	0.43	0.3	0.73
Doesn't like chocolates	0.08	0.19	0.27
Total	0.51	0.49	1

There are 27% that neither boys or girls like chocolates.

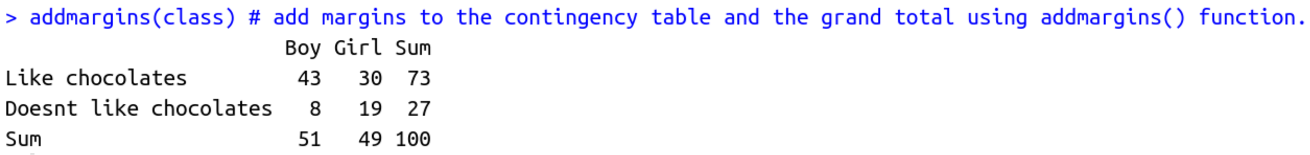
IST772 Summary Template: Chapter 2 – Reasoning with Probability

Originality Assertion: By submitting this file you affirm that this writing is your own.

Box on p.30: Make Your Own Tables with R

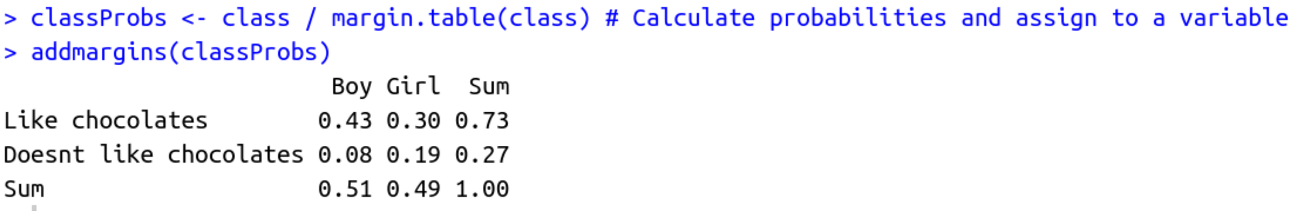


```
# In the example below, I have created a 2 by 2 matrix, by using matrix() function.
# In a primary school class, the teacher asked the boys and the girls weather they liked chocolates or not.

class <- matrix(c(43,30,8,19),ncol=2,byrow=TRUE)

colnames(class) <- c('Boy','Girl') # Label the columns
rownames(class) <- c('Like chocolates','Doesnt like chocolates') # Label the rows

class <- as.table(class) # convert to a table
class # show the table

addmargins(class) # add margins to the contingency table and the grand total using addmargins() function.
```



```
> addmargins(class) # add margins to the contingency table and the grand total using addmargins() function.
                       Boy Girl Sum
Like chocolates         43   30  73
Doesnt like chocolates   8   19  27
Sum                     51   49 100
```

to find the probabilities for each cell in the contingency table, we type the code below in R.



```
> classProbs <- class / margin.table(class) # Calculate probabilities and assign to a variable
> addmargins(classProbs)
                        Boy Girl  Sum
Like chocolates        0.43 0.30 0.73
Doesnt like chocolates 0.08 0.19 0.27
Sum                    0.51 0.49 1.00
```