605-HW06-Combinatorics

Michael Y.

October 6, 2019

Contents

HW6 - Combinatorics

Problem 1 - Marbles

A box contains 54 red marbles, 9 white marbles, and 75 blue marbles.

If a marble is randomly selected from the box, what is the probability that it is red or blue?

Express your answer as a fraction or a decimal number rounded to four decimal places.

```r
red = 54
white = 9
blue = 75
success01 = red+blue
total01 = red+white+blue
prob01 = success01/total01
roundprob01 = round(prob01,4)
roundprob01
```

```
## [1] 0.9348
```

$$P(Red \cup Blue) = \frac{Num(Red) + Num(Blue)}{Num(Red) + Num(White) + Num(Blue)} = \frac{54 + 75}{54 + 9 + 75} = \frac{43}{46} = 0.9348$$

Problem 2 - Minigolf

You are going to play mini golf. A ball machine that contains

19 green golf balls,

20 red golf balls,

24 blue golf balls, and

17 yellow golf balls, randomly gives you your ball.

What is the probability that you end up with a red golf ball?

Express your answer as a simplified fraction or a decimal rounded to four decimal places.

```
green = 19
red = 20
blue = 24
yellow = 17
success02 = red
total02 = green+red+blue+yellow
prob02 = success02/total02
prob02
```

```
## [1] 0.25
```

$$P(R) = \frac{Num(Red)}{Num(Green) + Num(Red) + Num(Blue) + Num(Yellow)} = \frac{20}{19 + 20 + 24 + 17} = \frac{1}{4} = 0.25$$

Table 1: **Gender and Residence of Customers**

	Males	Females
Apartment	81	228
Dorm	116	79
With Parent(s)	215	252
Sorority/Fraternity House	130	97
Other	129	72

Problem 3 - Pizza Delivery

A pizza delivery company classifies its customers by gender and location of residence.

The research department has gathered data from a random sample of 1399 customers.

The data is summarized in the table below.

```
matr03  %>% kable(caption="**Gender and Residence of Customers**")  %>%
          kable_styling(c("striped", "bordered"))
```

What is the probability that a customer is *not male* or *does not live with parents* ?
Write your answer as a fraction or a decimal number rounded to four decimal places.

By **De Morgan's law,** $(\neg Male) \vee (\neg WithParents) = \neg (Male \wedge WithParents)$

So:

```
totl = sum(matr03)
Males=sum(matr03[,"Males"])
notMales=totl - Males
WithParents = sum(matr03["With Parent(s)",])
notWithParents = totl - WithParents
MaleAndWithParents = matr03["With Parent(s)","Males"]
notMaleOrNotWithParents = totl - MaleAndWithParents
result = round(notMaleOrNotWithParents/totl,4)
```

```
## Total:                           1399
```

```
## Total Males:                  671

## Not Males:                    728

## With Parents:                 467

## Not With Parents:             932

## Male AND With Parents:        215

## NotMale OR NotWithParents:    1184

## result:                       0.8463
```

$$Pr[(\neg Male) \vee (\neg WithParents)] = 1 - Pr[(Male \wedge WithParents)] = 1 - \frac{215}{1399} = \frac{1184}{1399} = 0.8463$$

Problem 4 - Independence

Determine if the following events are independent.

Going to the gym.

Losing weight.

By Bayes Rule,

$$Pr(A|B) = \frac{Pr(A \wedge B)}{Pr(B)}$$

.
If the events are independent,

$$Pr(A \wedge B) = Pr(A) \cdot Pr(B)$$

, so for independent events,

$$Pr(A|B) = \frac{Pr(A \wedge B)}{Pr(B)} = \frac{Pr(A) \cdot Pr(B)}{Pr(B)} = Pr(A)$$

To be independent,

$$Pr(GoingtoGym \wedge LosingWeight) = Pr(GoingtoGym) \cdot Pr(LosingWeight)$$

Alternatively, the events are independent if

$$Pr(LosingWeight|GoingtoGym) = Pr(LosingWeight)$$

However, a person is more likely to lose weight if he/she does go to the gyn, i.e.

$$Pr(LosingWeight|GoingtoGym) > Pr(LosingWeight)$$

Therefore, these events are NOT independent.

Answer: ***A) Dependent*** B) ~~Independent~~

Problem 5 - Veggie Wrap

A veggie wrap at City Subs is composed of 3 different vegetables and 3 different condiments wrapped up in a tortilla.

If there are 8 vegetables, 7 condiments, and 3 types of tortilla available, how many different veggie wraps can be made?

```
vegetablechoice = choose(8,3)
condimentchoice = choose(7,3)
tortillachoice  = choose(3,1)
result = vegetablechoice * condimentchoice * tortillachoice
result
```

```
## [1] 5880
```

There are 56 ways to choose 3 vegetables from 8 and there are 35 ways to choose 3 condiments from 7.

Therefore the number of possible veggie wraps is $56 \cdot 35 \cdot 3 = 5880$.

Problem 6 - More indepedence

Determine if the following events are independent.

Jeff runs out of gas on the way to work.

Liz watches the evening news.

There does not appear to be any relationship between these two events, especially as they are involving separate people.

Answer: A) ~~Dependent~~ ***B) Independent***

Problem 7 - Cabinet

The newly elected president needs to decide the remaining 8 spots available in the cabinet he/she is appointing.

If there are 14 eligible candidates for these positions (where rank matters), how many different ways can the members of the cabinet be appointed?

Because order matters, we need to use permutations, $P(n,k) = \frac{n!}{(n-k)!}$, rather than combinations, $C(n,k) = \frac{n!}{k!(n-k)!}$.

```
permute <- function(n,k) {
  factorial(n) / factorial(n-k)
  }
candidates = 14
slots = 8
result = permute(candidates,slots)
#result
fmtresult = format(result,big.mark=",", trim=TRUE)
fmtresult
```

```
## [1] "121,080,960"
```

The number of permutations to fill 8 from 14, where sequence matters, is

$$\frac{14!}{(14-8)!} = \frac{14!}{(6)!} = 121,080,960$$

Problem 8 - Jellybeans

A bag contains 9 red, 4 orange, and 9 green jellybeans.

What is the probability of reaching into the bag and randomly withdrawing 4 jellybeans such that

the number of red ones is 0,

the number of orange ones is 1,

and the number of green ones is 3?

Write your answer as a fraction or a decimal number rounded to four decimal places.

```
totalRED = 9
totalORANGE = 4
totalGREEN = 9
totalJELLYBEANS = totalRED + totalORANGE + totalGREEN
chooseRED = 0
chooseORANGE = 1
chooseGREEN = 3
chooseJELLYBEANS = chooseRED + chooseORANGE + chooseGREEN

combRED = choose(totalRED,chooseRED)
combORANGE = choose(totalORANGE,chooseORANGE)
combGREEN = choose(totalGREEN,chooseGREEN)

combSELECTED = combRED * combORANGE * combGREEN
combJELLYBEANS = choose(totalJELLYBEANS,chooseJELLYBEANS)
result = round(combSELECTED / combJELLYBEANS,4)
result
```

```
## [1] 0.0459
```

There is 1 way to choose 0 Red jellybeans from 9 (i.e., choose none of them.)

There are 4 ways to choose 1 Orange jellybeans from 4.

There are 84 ways to choose 3 Green jellybeans from 9.

Thus, the number of ways to select the above combination is $1 \cdot 4 \cdot 84 = 336$.

There are 7315 ways to choose 4 jellybeans from 22 total jellybeans.

Therefore the probability of selecting the desired combination is $\frac{336}{7315} = 0.0459$.

Problem 9 - Factorials

Evaluate the following expression: $\frac{11!}{7!}$

```
result = factorial(11) / factorial(7)
result
```

```
## [1] 7920
```

The answer is $\frac{11!}{7!} = \frac{11 \cdot 10 \cdot 9 \cdot 8 \cdot 7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{7 \cdot 6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1} = 11 \cdot 10 \cdot 9 \cdot 8 = 7920$.

Problem 10 - Complement

Describe the complement of the given event:

"67% of subscribers to a fitness magazine are over the age of 34."

The **event** is $Age(Subscriber) > 34$,
and the **probability** of the event is $Pr(Age(Subscriber) > 34) = 0.67$.

In other words: choose a random subscriber to the fitness magazine.
The probability that the randomly selected individual is older than 34 is 0.67 .

The complement of the **event** is $Age(Subscriber) \leq 34$,
and the **probability** of the complement is $Pr(Age(Subscriber) \leq 34) = 1 - Pr(Age(Subscriber) > 34) = 0.67 = 0.33$.

In other words, the probability that a randomly selected subscriber to the magazine is age 34 or younger is 0.33 .

Problem 11 - Coin toss

If you throw exactly three heads in four tosses of a coin you win \$97. If not, you pay me \$30.

Step 1. Find the expected value of the proposition. Round your answer to two decimal places.

There are four ways to get exactly three heads in 4 tosses: {HHHT,HHTH,HTHH,THHH}.
The number of possible results in 4 tosses is $2^4 = 16$. Therefore, the probability that I win 97 dollars is $Pr(Win) = \frac{4}{16} = \frac{1}{4} = 0.25$, and the probability that I lose 30 dollars is $Pr(Lose) = 1 - Pr(Win) = 1 - 0.25 = 0.75$.

```
winget = 97
losspay = -30
waystowin = choose(4,3)
eventspace = 2^4
waystolose = eventspace - waystowin
probwin = waystowin / eventspace
#probwin
problose = waystolose / eventspace
#problose
expected = round(probwin * winget + problose * losspay,2)
#expected
```

The expected value of the game is

$$Pr(Win) \cdot Payoff(Win) + Pr(Lose) \cdot Payoff(Lose) = 0.25 \cdot \$97 - 0.75 \cdot \$30 = 1.75$$

Step 2. If you played this game 559 times how much would you expect to win or lose? (Losses must be entered as negative.)

```
plays = 559
cumulative = plays * expected
#cumulative
```

The expected result of playing 559 times is $\$1.75 \cdot 559 = \$978.25$.

Problem 12 - More coin toss

12. Flip a coin 9 times. If you get 4 tails or less, I will pay you $23. Otherwise you pay me $26.

Step 1. Find the expected value of the proposition. Round your answer to two decimal places.

```
successtails = (0:4)
successes = array(choose(9,successtails),dimnames=list(paste(successtails,"tail(s)")))
#successes
numsuccesses = sum(successes)
#numsuccesses
totalpossibilities = 2^9
#totalpossibilities
probsuccess = numsuccesses / totalpossibilities
#probsuccess
probfailure = 1 - probsuccess
#probfailure
winget = 23
losepay = -26
expectedgame = winget * probsuccess + losepay*probfailure
#expectedgame
```

The number of ways to get between 0 and 4 tails in 9 tosses is

0 tail(s)	1 tail(s)	2 tail(s)	3 tail(s)	4 tail(s)
1	9	36	84	126

and the sum of this table is 256. The total number of possible results from 9 tosses is $2^9 = 512$,

so the probability of winning is $\frac{256}{512} = 0.5$ and the probability of losing is $1 - 0.5 = 0.5$.

The expected result from one play of this game is thus

$$0.5 \cdot 23 \text{ dollars} + 0.5 \cdot (-26) \text{ dollars} = -1.5 \text{ dollars}$$

.

Step 2. If you played this game 994 times how much would you expect to win or lose? (Losses must be entered as negative.)

```
plays = 994
cumulative = plays * expectedgame
#cumulative
```

If this game were played 994 times, the expected result would be $-1491 dollars, i.e., you would expect to lose $1491 .

Problem 13 - Polygraph

13. The sensitivity and specificity of the polygraph has been a subject of study and debate for years.

A 2001 study of the use of polygraph for screening purposes suggested that:

the probability of detecting a liar was .59 (sensitivity) and that

the probability of detecting a "truth teller" was .90 (specificity).

We estimate that about 20% of individuals selected for the screening polygraph will lie.

```
sensitivity = 0.59
specificity = 0.90
liars = 0.20
truthtellers = 1-liars
```

In the analysis, "Positive" indicates a liar, while "Negative" indicates a Truthteller.

```
grid = matrix(c("TP","FP","TP+FP",
                "FN","TN","FN+TN",
                "TP+FN","FP+TN","TP+FN+FP+FN"),
              3,3,byrow=T,
              dimnames = list(c("PredictedLiar","PredictedTruthTeller","TotalPredicted"),
                              c("ActualLiar","ActualTruthTeller","TotalActual")))
grid %>% kable() %>% kable_styling(c("striped", "bordered"))
```

	ActualLiar	ActualTruthTeller	TotalActual
PredictedLiar	TP	FP	TP+FP
PredictedTruthTeller	FN	TN	FN+TN
TotalPredicted	TP+FN	FP+TN	TP+FN+FP+FN

Since 20 percent of the individuals are actually Liars, $TP + FN = 0.2$, which means $FP + TN = 1 - 0.2 = 0.8$.

```r
grid["TotalPredicted","ActualLiar"]         = paste(grid["TotalPredicted",
                                                    "ActualLiar"],"=0.200")
grid["TotalPredicted","ActualTruthTeller"] = paste(grid["TotalPredicted",
                                                    "ActualTruthTeller"],"=0.800")
grid["TotalPredicted","TotalActual"]        = paste(grid["TotalPredicted",
                                                    "TotalActual"],"=1.000")
grid %>% kable() %>% kable_styling(c("striped", "bordered"))
```

	ActualLiar	ActualTruthTeller	TotalActual
PredictedLiar	TP	FP	TP+FP
PredictedTruthTeller	FN	TN	FN+TN
TotalPredicted	TP+FN =0.200	FP+TN =0.800	TP+FN+FP+FN =1.000

$$
\begin{aligned}
\text{Sensitivity} &= \frac{\text{TruePositives}}{\text{TruePositives+FalseNegatives}} \\
&= \frac{TP}{TP+FN} = \frac{TP}{0.200} = 0.59 \\
&= \frac{\text{ActualLiars}}{\text{ActualLiars+MislabeledTruthtellers}}
\end{aligned}
$$

Therefore, $TP = (TP + FN) \cdot 0.59 = 0.200 \cdot 0.59 = 0.118$,
and $FN = (TP + FN) - TP = 0.200 - 0.118 = 0.082$.

```r
grid["PredictedLiar","ActualLiar"]          = paste(grid["PredictedLiar",
                                                    "ActualLiar"],"=0.118")
TP=0.118
grid["PredictedTruthTeller","ActualLiar"]  = paste(grid["PredictedTruthTeller",
                                                    "ActualLiar"],"=0.082")
FN=0.082
grid %>% kable() %>% kable_styling(c("striped", "bordered"))
```

	ActualLiar	ActualTruthTeller	TotalActual
PredictedLiar	TP =0.118	FP	TP+FP
PredictedTruthTeller	FN =0.082	TN	FN+TN
TotalPredicted	TP+FN =0.200	FP+TN =0.800	TP+FN+FP+FN =1.000

$$
\begin{aligned}
\text{Specificity} &= \frac{\text{TrueNegatives}}{\text{FalsePositives+TrueNegatives}} \\
&= \frac{TN}{FP+TN} = \frac{TN}{0.800} = 0.90 \\
&= \frac{\text{ActualTruthTellers}}{\text{MislabeledLiars+ActualTruthtellers}}
\end{aligned}
$$

Therefore, $TN = (FP + TN) \cdot 0.90 = 0.800 \cdot 0.90 = 0.720$,
and $FP = (FP + TN) - TN = 0.800 - 0.720 = 0.080$.

```
grid["PredictedLiar","ActualTruthTeller"]  = paste(grid["PredictedLiar",
                                                        "ActualTruthTeller"],"=0.080")
FP=0.080
grid["PredictedTruthTeller","ActualTruthTeller"]  = paste(grid["PredictedTruthTeller",
                                                        "ActualTruthTeller"],"=0.720")
TN=0.720
grid %>% kable() %>% kable_styling(c("striped", "bordered"))
```

	ActualLiar	ActualTruthTeller	TotalActual
PredictedLiar	TP =0.118	FP =0.080	TP+FP
PredictedTruthTeller	FN =0.082	TN =0.720	FN+TN
TotalPredicted	TP+FN =0.200	FP+TN =0.800	TP+FN+FP+FN =1.000

Completing the rightmost totals in the grid:

```
grid["PredictedLiar","TotalActual"]  = paste(grid["PredictedLiar",
                                                  "TotalActual"],"=0.198")
grid["PredictedTruthTeller","TotalActual"]  = paste(grid["PredictedTruthTeller",
                                                  "TotalActual"],"=0.802")
grid %>% kable() %>% kable_styling(c("striped", "bordered"))
```

	ActualLiar	ActualTruthTeller	TotalActual
PredictedLiar	TP =0.118	FP =0.080	TP+FP =0.198
PredictedTruthTeller	FN =0.082	TN =0.720	FN+TN =0.802
TotalPredicted	TP+FN =0.200	FP+TN =0.800	TP+FN+FP+FN =1.000

Confirm that these results give the correct metrics:

```
Check_Sensitivity = TP / (TP+FN)
Check_Sensitivity
```

```
## [1] 0.59
```

```
sensitivity == Check_Sensitivity
```

```
## [1] TRUE
```

```
Check_Specificity = TN / (FP+TN)
Check_Specificity
```

```
## [1] 0.9
```

```
specificity == Check_Specificity
```

```
## [1] TRUE
```

a. What is the probability that an individual is actually a liar given that the polygraph detected him/her as such? (Show me the table or the formulaic solution or both.)

The percentage which are actually liars is TP =0.118 .

The percentage predicted to be liars is TP+FP =0.198 .

```
correctlyPredictedLiars=round(TP/(TP+FP),4)
correctlyPredictedLiars
```

```
## [1] 0.596
```

The probability that an individual is ***actually*** a liar, given that the polygraph detected him/her as such is 0.596 .

This metric is known as **precision** or **positive predictive value.**

b. What is the probability that an individual is actually a truth-teller given that the polygraph detected him/her as such? (Show me the table or the formulaic solution or both.)

The percentage which are actually truthtellers is TN =0.720 .

The percentage predicted to be truthtellers is FN+TN =0.802 .

```
correctlyPredictedTruthTellers=round(TN/(FN+TN),4)
correctlyPredictedTruthTellers
```

```
## [1] 0.8978
```

The probability that an individual is ***actually*** a truth-teller, given that the polygraph detected him/her as such, is 0.8978.

This metric is known as **negative predictive value.**

c. What is the probability that a randomly selected individual is either a liar or was identified as a liar by the polygraph? Be sure to write the probability statement.

By DeMorgan's Law,

$$\begin{aligned}
Pr(ActualLiar \vee PredictedLiar) &= 1 - Pr(\neg(ActualLiar) \wedge \neg(PredictedLiar) \\
&= 1 - Pr(ActualTruthTeller \wedge PredictedTruthTeller) \\
&= 1 - 0.72 \\
&= 0.28
\end{aligned}$$