GR5065 Homework 1

Cindy Li

Due February 7, 2022 at 6PM

1 Poker

1.1 Probability of rare five-card combinations

```
## probability of a flush (no straight flush & royal flush)
dhyper(x = 5, m = 13, n = 52 - 13, k = 2 + 5) * 4
```

```
## [1] 0.02851351
```

The probability of ending up with a 'flush' is given above. Because we have m = 13 cards in each suit, and we need to draw x = 5 of them for a flush, there are n = 39 cards left in the deck, and you have k = 7 opportunities to draw 5 cards in the same suit. There are 4 suits so we multiply this probability by 4.

```
## probability of a full house - probability of 4 of a kind

P_three <- dhyper(x = 3, m = 4, n = 52 - 4, k = 5 + 2) * 13 #P(A)
P_pairgiventhree <- dhyper(x = 2, m = 4, n = 49 - 4, k = 4) * 12 #P(B|A)
P_pair <- dhyper(x = 2, m = 4, n = 52 - 4, k = 5 + 2) * 13 #P(B)

P_threegivenpair <- (P_three * P_pairgiventhree) / P_pair #P(A|B)

P_four <- dhyper(x = 4, m = 4, n = 52 - 4, k = 2 + 5) * 13 #P(X)
P_fourgivenfull <- dhyper(x = 4, m = 4, n = 52 - 5, k = 7) * 2 # P(Y|X) where P(Y) = P(A|B)
P_fullgivenfour <- P_four * P_fourgivenfull / P_threegivenpair #P(X|Y)

P_threegivenpair
```

```
## [1] 0.02548661
```

```
P_fullhouse <- P_threegivenpair - P_fullgivenfour
P_fullhouse
```

```
## [1] 0.02546813
```

The probability of a 'full-house' is given above. We want to get a 3-of-a-kind, and then 2-of-a-kind of a different value, with neither being a 4-of-a-kind.

The first part of the equation gives us the probability of 3-of-a-kind (let's call this A), where we need to draw x = 3 from m = 4 cards, so there are n = 48 cards left in the deck, with k = 7 opportunities to get 3-of-a-kind. We then multiply this by 13 for each value in the deck - this gives P(A).

The second part of the equation gives the probability of a pair (B) GIVEN A. So again, you need to draw x = 2 cards from m = 4 cards, and since we have 8 'good cards' we can potentially draw, n = 44. Given A (7-3), we only have k = 4 opportunities to get a pair - B, and given that we have already claimed 1/13 values for A, we multiply this by 12 for the rest of the card values in the deck.

The third part of the equation divides the above multiplication by P(B) - the probability of getting a pair, where x = 2, m = 4, n = 48, and k = 7, multiplied by 13, similar to how we got P(A).

Finally, we want to subtract the probability that one of these two values (A or B) are a 4-of-a-kind. This can be calculated by looking at the conditional probability of four-of-a-kind given that there is a full house, similar to the above.

```
#probability of four-of-a-kind
dhyper(x = 4, m = 4, n = 52 - 4, k = 2 + 5) * 13
```

```
## [1] 0.001680672
```

```
## probability of a full house - probability of 4 of a kind
P_fullhouse
```

```
## [1] 0.02546813
```

```
## probability of a flush (no straight flush & royal flush)
dhyper(x = 5, m = 13, n = 52 - 13, k = 2 + 5) * 4
```

```
## [1] 0.02851351
```

```
#probability of three-of-a-kind
dhyper(x = 3, m = 4, n = 52 - 4, k = 2 + 5) * 13
```

```
## [1] 0.07563025
```

Probabilities of key five-card combinations are given above.

1.2 Pre-flop action

```
## probability of pocket aces
dhyper(x = 2, m = 4, n = 52 - 4, k = 2)
```

```
## [1] 0.004524887
```

From Selbst's perspective, the probability of having two Aces = the probability of being dealt two Aces because she definitely already has two Aces (pocket aces!). So P(AAdealt) = P(AAhave) = 1.

From Baumann's perspective, the probability of Selbst's being dealt two Aces is higher than the probability of having two Aces. This is because Baumann has two cards that are not Aces, meaning the probability of

Selbst being dealt 2 Aces increases from 4/52 to 4/50. Since Baumann has the new information that her two cards are NOT Aces, this changes her perspective of the probability of someone else having Aces - so P(AAdealt) > P(AAhave)

From Rosen's perspective, the probability that Selbst has two Aces is lower than the probability of having two Aces. Similar to the above, Rosen has new information, where he has one card that is an Ace. So this changes his perspective of Selbst's probability of being dealt 2 Aces to 3/50 from 4/52, thus decreasing Selbst's probability of being dealt two Aces. So in his perspective, P(AAdealt) < P(AAhave), again, since he holds one of 4 Aces in his hands (hole cards).

1.3 Schwartz's call

Despite having a less than average hand, Schwartz still makes the right decision calling because of the size of the pot (1025) is much greater than the 250 additional chips he would have to put in to get potential realized value - 250/1025 is around 25%, which is a very high "pot odds". In this sense, the 25% investment in the pot is greater than the probability of most above average hands of winning. As such, the expected value from his decision to call is greater than the risk, at least at this point, assuming he does not know what the other two players have.

Further, assuming the other two players both have "good" hands, they are looking for 5-card combinations in the same range. Given that this is a multi-player game, this means that they are potentially 'blocking' each other (i.e. one player has a card that another player needs in their hand). Schwartz, who has a hole cards that are below average, may have an opportunity to hit something good in the flop - since the probability of each card coming out of the flop/turn/river are the same for everyone. Since his cost to see the flop is so much smaller, relative to the current size of the pot, it makes sense for him to call at 400 chips.

1.4 Flop action

If Schwartz's utility function were equal to his chips at the conclusion of the hand and Schwartz would be indifferent between folding and calling Selbst's bet of 700 chips (which Baumann called, bringing the pot to 2, 675 chips) if he had the Q and 8 , what must the conditional probability be that either Selbst or Baumann has the K ?

If Schwartz would be indifferent between folding and calling if he had the Q of clubs, it means that he would only care if Selbst or Baumann had the K of clubs. We also assume here that a K-flush would win, regardless of what happens after (obviously quad 7s wins). So all we are looking for here is the conditional probability that Selbst or Baumann have a K of clubs, given what Schwartz already knows.

p(Q) = 1 Essentially - this is 47 choose 1 * 2 (for 2 other players) - I do not believe we actually need to expand the whole Baye's rule equation here, since we already know the premise of the flop without necessarily needing the marginal probabilities of the flop and/or just drawing the K of clubs.

```r
P_Q <- dhyper(x = 1, m = 1, n = 52 - 1, k = 2)
P_K <- dhyper(x = 1, m = 1, n = 52 - 1, k = 2)

P_Qgivenflop <- dhyper(x = 1, m = 1, n = 47 - 1, k = 4)

P_Kgivenflop <- P_Qgivenflop * P_K / P_Q

P_Kgivenflop
```

```
## [1] 0.08510638
```

```r
#this is the same as 47 choose 1 for 2 people - since we already know the other side, we don't have to 

P_Kclub <- dhyper(x = 1, m = 1, n = 47 - 1, k = 2) * 2
P_Kclub
```

```
## [1] 0.08510638
```

The conditional probability that one of Baumann or Selbst has the K of clubs is ~8.5%.

1.5 Pre-turn calculation

By this point, Baumann would only win if one of the two last cards are a 7, and they would tie if the last two cards drawn are both clubs, and Selbst would win in all other cases.

As such, we look at the probability of the turn or river being the 7 of spades, where x = 1, drawing from only m = 1 'good card', from 45 cards in the deck less the 1 good card n = 44, with k = 2 opportunities.

Similarly, a flush is only possible for both players if the last x = 2 cards are clubs, from 13 - 3 (since 3 are in the flop) so m = 10 clubs. There are 45 - 10 'bad cards' left in the deck, and k = 2 opportunties to draw those clubs.

Finally, one of these events must happen, so we subtract the probability of those two events from 1, giving the probability that Baumann wins. The calculations for these are shown below - which gives us the same percentages as the video.

```r
## we assume we do not count Schwartz' cards since he has folded

P_7 <- dhyper(x = 1, m = 1, n = 45 - 1, k = 2) #Baumann wins
P_7
```

```
## [1] 0.04444444
```

```r
P_clubs <- dhyper(x = 2, m = 13 - 3, n = 45 - 10, k = 2) #tie
P_clubs
```

```
## [1] 0.04545455
```

```r
P_Swin <- 1 - P_7 - P_clubs #Selbst wins
P_Swin
```

```
## [1] 0.910101
```

1.6 Turn

In the current state of the hands, the probability of one player ending up with a full house and the other ending up with a four-of-a-kind is the joint probability of the dealer dealing an Ace or a 7 (which is what happened), given what is already in the community. However, since a player already folded with an Ace in their hand, the probability of 4 aces is 0, so the probability at this point is only the probability of drawing a 7 - which we've seen above is 0.0444. The other probability is given below.

```r
P_7 <- dhyper(x = 1, m = 1, n = 45 - 1, k = 2) #probability of 7
P_A <- dhyper(x = 1, m = 1, n = 45 - 1, k = 2) #probability of A

P_FHOr4 <- P_7 + P_A

P_FHOr4
```

```
## [1] 0.08888889
```

If we consider before the hand starts at all, the probability of a player having a full house (A) and another player ending up with a four-of-a-kind (B) is the probability of getting a full house given we have four-of-a-kind P(B|A).

P(B|A) = P(B) * P(A|B) / P(A)

```r
P_A = P_threegivenpair #from Q1

P_four <- dhyper(x = 4, m = 4, n = 52 - 4, k = 2 + 5) * 13 #P(B)
P_fourgivenfull <- dhyper(x = 4, m = 4, n = 52 - 5, k = 7) * 2 # P(A|B) where P(A)
P_fullgivenfour <- P_four * P_fourgivenfull / P_A #P(B|A)
P_fullgivenfour
```

```
## [1] 1.847153e-05
```

We see above that it is extremely unlikely for a player to have a full-house and another to have four-of-a-kind in the same hand.

1.7 River action

In Selbst's perspective, given what she knows, the only two potential hands that Baumann could have are A/7, in which case Selbsts wins, or 7/7, in which case she loses. In saying so, in Selbst's perspective, she essentially has a 1/2 chance of winning, given what she knows about what is on the board and betting strategies. The small inconsistency here is that had Baumann had A/7, she likely would have folded earlier.

In terms of Bayes' rule, before considering any additional information on the table, Baumann could have any 2 cards. However, using Bayes' rule, we take into account what is on the table, and the betting actions. Baumman's bet suggests she could have one of 2 hands. Our calculations below shows that the probability of Baumann having A/7 is actually higher than the probability of her having pocket 7s. In this case, not considering betting behaviors, her probability of winning is higher, since she has the Ace full with 7s, which would win. What makes her decision more difficult is the fact that Baumman could have folded earlier if he had A/7 - but without taking this into consideration, the probability of winning for Selbst is higher - therefore it makes sense for her to call Baumann's raise.

Let P(7GF) = probability of hole card 7 given the flop P(AceGF) = probability of hole card A given the flop P(27FG) = probability of pocket 7s P(Ace7) = probability of hole cards Ace and 7

```r
P_7GF <- dhyper(x = 1, m = 2, n = 49 - 1, k = 2)
P_AceGF <- dhyper(x = 1, m = 1, n = 49 - 1, k = 2)
P_27GF <- dhyper(x = 2, m = 2, n = 49 - 1, k = 2)

P_Ace7 = P_7GF * P_AceGF

P_Ace7
```

```
## [1] 0.003198667
```

```r
P_27GF
```

```
## [1] 0.0008163265
```

```r
##note we do not use bayes rule more formally here because we can show the same probabilities without e
```

From the above, we see that P__Ace7 > P__27GF - thus confirming the explanation above.

1.8 Baumann's strategy on the river

Generally, if Selbst knows that Baumann will only shove (all-in) with pair Aces or pair sevens, Baumann's expected value would be lower (than if she were to bluff sometimes).

Assuming that 1) we see an Ace and 7 in the flop, and 2) the probability of the hands we are discussing are uniform, as Baumann, there are three combinations of Aces possible that may win (quad, full house, ace high pair), and 1 combination of 7s possible that would win (quad 7s). So, if Baumann has pocket 7s, 3/4 of the time she may lose - and only has a 25% chance of winning. In this case, even if she all-ins, she may not realize the full value. In the case where she has pocket Aces, if she all-ins, Selbst could have the 7s, and therefore she would lose her bet.

In either case, if she never bluffs, not only is there no additional value to raising all-in (since if Selbsts has a worse hand, she will just fold), she risks losing value as well (in the cases where Selbsts has a better hand).

As an added note, if Baumann occasionally bluffs (i.e. Selbsts does not have perfect information of her opponent's strategy), the Baumann's chip expectation is higher.

2 Surnames

2.1 Notation

$$\boldsymbol{Pr\left(\cancel{r_i} \cap r_j = r \,\middle|\, S_i = s, G_i = g\right) = \frac{Pr(G_i = g \mid \cancel{r_i} \cap r_j = r)\, Pr(\cancel{r_i} \cap r_j = r \mid S_i = s)}{\sum_{r' \in R} Pr\,(G_i = g \mid \cancel{r_i} \cap r_j = r')\, Pr(\cancel{r_i} \cap r_j = r' \mid S_i = s)}}$$

Figure 1: notation

2.2 Frequentist Perspective

Frequentists such as Fisher would not approve of this use of Bayes' Rule in the paper - because they don't agree with Bayes' Rule generally. More specifically, the idea that a person can be "98% likely" to be something is inconceivable. Now - I'm not saying a person can't be 50% one race and 50% another (mixed babies exist), however, the idea that you would change your belief about someone's race based on data - someone is either X race or they are not - and that is a factor of who they are, and not the data we have gathered, and data cannot change who they are substantively.

However, the way Bayesian statisticians look at it, as we find out more information i.e. geography, sex, last name, age, etc. about someone, we become more or less certain that they are of a certain rate.

This is not to say that certain methods in the paper are not used by frequentists i.e. confusion matrices, ROC, etc., but it is the general idea that challenges the ideals of a frequentist. Therefore, the fact that we are using probabilities to describe someone with a certain (in frequentist views) race that is impossible to change would drive Fisher mad.

2.3 You

```r
## load data
NYS <- readRDS("NYS.rds")

## commented the below out to keep assignment shorter
#library(dplyr)
#select(NYS$NY$county, county, county_name) %>%
#arrange(county_name)
```

From the table: County = New York County, 061

```r
## create dataframe

surname <- "Li"
state <- "NY"
county <- "061"
age <- 24
sex <- 1

my_df <- data.frame(surname, state, county, age, sex)
```

Created my dataframe and read in the data

```r
library(wru)
me <- predict_race(my_df, census.geo = "county", census.data = NYS, age = TRUE, sex = TRUE)
```

```
## [1] "Proceeding with Census geographic data at county level..."
## [1] "Using Census geographic data from provided census.data object..."
## [1] "State 1 of 1: NY"
```

```r
me
```

```
##   surname state county age sex    pred.whi     pred.bla    pred.his pred.asi
## 1      Li    NY    061  24   1 0.008135272 0.0007777129 0.004087571 0.980841
##      pred.oth
## 1 0.006158488
```

The mode above predicts (with almost certainty) that I am of Asian decent - which is correct!

2.4 Granularity

Increasing the granularity (and thus decreasing the denominator) should increase the lower probabilities and decrease the larger probabilities (will lower the amount of predictive power generally) since you are decreasing your sample. Thus, if we ONLY used the tract-level data, with a frequentist approach, increasing the granularity may decrease the precision of the probabilities. However, since we are using Bayes, and we use the county level data as a prior, and use our tract data as new data that updates our existing numbers, we may improve the accuracy (decreasing false positives & negatives) of the individual predictions.