18.05 Problem Set 1, Spring 2022 Solutions

Problem 1. (20 pts.) **(6 card draw)**
You can easily look up the probability of 5 card poker hands, e.g. **https://en.wikipedia.org/wiki/Poker_probability**

For this problem, let's consider hands with 6 cards. Here are two types:

Two-pair*: Two cards have one rank, two cards have another rank, and the remaining two cards have two different ranks. e.g.* $\{2\heartsuit, 2\spadesuit, 5\heartsuit, 5\clubsuit, Q\diamondsuit, K\diamondsuit\}$

Three-of-a-kind*: Three cards have one rank and the remaining three cards have three other ranks. e.g.* $\{2\heartsuit, 2\spadesuit, 2\clubsuit, 5\clubsuit, 9\spadesuit, K\heartsuit\}$

Calculate the probability of each type of hand. Which is more probable?

Solution: (Reasons below.)
$P(\text{two-pair}) = 0.1214, \quad P(\text{three-of-a-kind}) = 0.03596,$ two pairs is more likely

We create each hand by a sequence of actions and use the rule of product to count how many ways it can be done. (Critically: the number of choices available at each step is independent of the choices made in the earlier steps.)

two pair
We build a hand with two pair with the following steps.

1. Choose 2 ranks –one for each pair: $\binom{13}{2}$ ways.
2. Choose 2 cards from each chosen rank: 4 cards per suit: $\binom{4}{2}\binom{4}{2}$ ways.
3. Choose 2 of the remaining ranks for the remaining 2 cards in the hand: $\binom{11}{2}$ ways.
4. Choose 1 card from each of these ranks: $\binom{4}{1}\binom{4}{1}$ ways

So, the number of hands that are two pair is

$$\binom{13}{2}\binom{4}{2}\binom{4}{2}\binom{11}{2}\binom{4}{1}\binom{4}{1} = 2471040.$$

We divide this by the total number of hands to get

$$P(\text{two pair}) = \frac{2471040}{\binom{52}{6}} \approx 0.1214,$$

Note that this gives the probability only because every hand is equally probable.

Three-of-a-kind:
We build a hand with three of a kind with the following steps.

1. Choose the rank of the triple: $\binom{13}{1}$ ways.
2. Choose 3 cards from the chosen rank: 4 cards per suit: $\binom{4}{3}$ ways.

3. Choose 3 of the remaining ranks for the remaining 3 cards in the hand: $\binom{12}{3}$ ways.

4. Choose 1 card from each of these ranks: $\binom{4}{1}\binom{4}{1}\binom{4}{1}$ ways

So, the number of hands that are three of a kind is

$$\binom{13}{1}\binom{4}{3}\binom{12}{3}\binom{4}{1}\binom{4}{1}\binom{4}{1} = 732160.$$

We divide this by the total number of hands to get

$$P(\text{two pair}) = \frac{732160}{\binom{52}{6}} \approx 0.03596,$$

Note that this gives the probability only because every hand is equally probable.

Problem 2. (20: 10,10 pts.) **(Non-transitive dice)**
In class we worked with non-transitive dice:

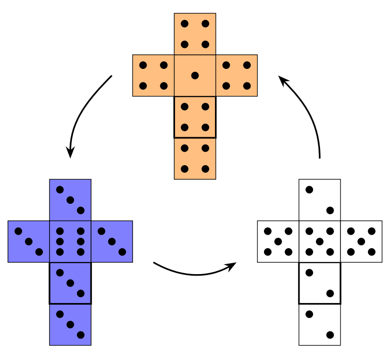
Blue: 3 3 3 3 3 6; Orange: 1 4 4 4 4 4; White: 2 2 2 5 5 5.

(a) *Find the probability that white beats orange, the probability that orange beats blue and the probability that blue beats white.*

Can you line the dice up in order from best to worst? (Hint: this is why these are called 'non-transitive'.)

Solution: (Reasons below.)
(a) P(white beats orange) = 7/12
P(orange beats blue) = 25/36
P(blue beats white) = 7/12.

No you can't line them up since blue beats white beats orange beats blue. You have to arrange them in a circle. This was the meaning of the graphic in the class 2 slides.



The reasons for these answers come from the probability tables.

	Blue die		White die		Orange die	
Outcomes	3	6	2	5	1	4
Probability	5/6	1/6	1/2	1/2	1/6	5/6

The 2×2 tables just below show the dice matched against each other. Each entry is the probability of seeing the pair of numbers corresponding to that entry. We color the probability with the color of the winning die for that pair. (For obvious reasons we use black instead of white when the white die wins.)

		White		Orange	
		2	5	1	4
Blue	3	5/12	5/12	5/36	25/36
	6	1/12	1/12	1/36	5/36
Orange	1	1/12	1/12		
	4	5/12	5/12		

Totaling the winning probabilities for each pair gives the probabilities stated at the start of the solution.

(b) *Suppose you roll two white dice against two blue dice. What is the probability that the sum of the white dice is greater than the sum of the blue dice?*

For hints on this problem and ways to make money with your dice, watch at least the first six minutes of the following video.

https://www.youtube.com/watch?v=zWUrwhaqq_c

Here a tree is used to organize the calculation rather than a table. We will discuss this method in Week 2.

Solution: P(sum of 2 white beats sum of 2 blue) = 85/144. Notice, this is a little surprising since a single blue tends to beat a single white.

The reasoning is similar to part (a) with slightly larger tables.

	2 Blue die			2 White die		
Outcomes	6	9	12	4	7	10
Probability	25/36	10/36	1/36	1/4	1/2	1/4

		2 White		
		4	7	10
	6	25/144	50/144	25/144
2 Blue	9	10/144	20/144	10/144
	12	1/144	2/144	1/144

Summing the different entries, We see, $P(\text{white wins}) = 85/144$ (so $P(\text{blue wins}) = 59/144$).

Problem 3. (55: 5,5,10,5,10,10,10 pts.) **(Birthdays: counting and simulation)**
Ignoring leap days, the days of the year can be numbered 1 to 365. Assume that birthdays are equally likely to fall on any day of the year. Consider a group of n people, of which you are not a member. An element of the sample space S will be a sequence of n birthdays (one for each person).

(a) *Define the probability function P for S. (This will depend on n.)*

Solution: The sample space S is the set of all sequences of n birthdays. That is, all sequences

$$\omega = (b_1, b_2, b_3, \ldots, b_n),$$

where each entry is a number between 1 and 365.

There are 365^n sequences of n birthdays. Since they are all equally likely, $P(\omega) = \frac{1}{365^n}$ for every sequence ω.

(b) *Consider the following events:*
A: "someone in the group shares your *birthday"*
B: "some two people in the group share a birthday"
C: "some three people in the group share a birthday"
Carefully describe the subset of S that corresponds to each event.

Solution: Event A: Suppose my birthday is on day b. Then "an outcome ω is in A" is equivalent to "b is in the sequence for ω", i.e. $b = b_k$ for some index k between 1 and n. More symbolically,

$$\text{an outcome } \omega \text{ is in } A \quad \text{if and only if} \quad b_k = b \text{ for some index } k \text{ in } 1, \ldots, n.$$

Event B: "An outcome ω is in B" is equivalent to "two of the entries in ω are the same". That is, an outcome ω is in B if and only if $b_j = b_k$ for two (different) indices j, k in $1, \ldots, n$.

Event C: an outcome ω is in C if and only if $b_j = b_k = b_l$ for three (distinct) indices j, k, l in $1, \ldots, n$.

(c) *Find an exact formula for $P(A)$. What is the smallest n such that $P(A) > 0.5$?*

Solution: It's easier to calculate $P(A^c)$. Since there are 364 birthdays that are not mine, there are 364^n outcomes in A^c.

$$P(A) = 1 - P(A^c) = 1 - \frac{364^n}{365^n}.$$

We can find the size of the group needed for $P(A) > 0.5$ by trial and error, plugging in different values of n. Or we can set $P(A) = 0.5$ and solve for n.

$$1 - \frac{364^n}{365^n} = 0.5 \quad \Rightarrow \quad \left(\frac{364}{365}\right)^n = 0.5 \quad \Rightarrow \quad n \cdot \ln\left(\frac{364}{365}\right) = \ln(0.5) \quad \Rightarrow \quad n \approx 252.65$$

So there needs to be at least 253 people for it to be more likely than not that one of them shares your birthday.

(d) *Justify why n in part (c) is greater than $\frac{365}{2}$ without doing any computation. (We are looking for a short answer giving a heuristic sense of why this is so.)*

Solution: Ignoring the fractions, 365/2 different birthdays would have a 50 percent chance of matching your birthday. But, 365/2 people probably don't all have different birthdays, so they have a less than 50 percent chance of matching.

(e) *Use R simulation to estimate the smallest n for which $P(B) > 0.9$. For these simulations, let the number of trials be 10000. (You can reuse your code from Studio 1.)*

For this value of n, repeat the simulation a few times to verify that it always gives similar results.

Using 10000 trials you saw very little variation in the estimate of $P(B)$. Try this again using 30 trials and verify that the estimated probabilities are much more variable. On your pset, give the results 7 runs of 30 trials using the value of n you just found.

Solution: Here's the R code I ran to estimate $P(B)$.

```
# colMatches is an 18.05 function. It needs to be in the R working directory
# You can find the file mit18_05_s22_colMatches.r on our class website.
source('mit18_05_s22_colMatches.r')
```

```
# Set up the parameters
ndays = 365
npeople = 20
ntrials = 10000
size_match = 2

year = 1:ndays

# Run ntrials -one per column- using sample() and matrix()
y = sample(year, npeople*ntrials, replace=TRUE)
trials = matrix(y, nrow=npeople, ncol=ntrials)
w = colMatches(trials, size_match)
prob_B = mean(w)
print(prob_B)
```

I ran the code with various values of `npeople`. Here is a table of the estimated values of $P(B)$.

`npeople`	$P(B)$ (multiple values means multiple runs of the code)
20	0.4123
30	0.7007
40	0.8913, 0.8867, 0.8926, 0.8865
41	0.9005, 0.9003, 0.9037, 0.9006, 0.903

We see that the estimated probability of B is consistently less than 0.9 for `npeople = 40` and consistently greater than 0.9 for `npeople = 41`. Therefore: **Answer:** 41.

When we run the code with `ntrials = 30` and `npeople = 41` we get the following estimates for $P(B)$.

$$0.9333, 0.9333, 0.9, 0.8333, 0.9333, 0.8, 1$$

This is certainly much more variable than the estimates in in the table above.

(f) *Find an exact formula for $P(B)$.*

Solution: It's easier to calculate $P(B^c)$, the probability that all n birthdays are distinct. To choose n different birthdays: there are 365 choices for the first birthday, 364 for the second birthday, etc. So

$$P(B) = 1 - P(B^c) = 1 - \frac{365 \cdot 364 \cdots (365 - n + 1)}{365^n} = 1 - \frac{365!}{(365 - n)! \cdot 365^n}.$$

(g) *Use R simulation to estimate the smallest n for which $P(C) > 0.5$. Again use 10000 trials. You will find that two adjacent values of n are equally plausible based on simulations. You may pick either one for your answer.*

Note that it is much harder to find an exact formula for $P(C)$, so simulation is especially handy.

Solution: To estimate $P(C)$ we used the same code as in part (e), except we set `size_match = 3`. Here is the table of (estimated) probabilities we found.

`npeople`	$P(C)$ (multiple values means multiple runs of the code)
20	0.0121
30	0.0293
50	0.1274
80	0.4241
90	0.5389
85	0.4832
86	0.4841
87	0.502, 0.4989, 0.4909, 0.5035
88	0.5081, 0.5115, 0.5149, 0.5071

The estimates for 87 are sometimes above and sometimes below 0.5. So the answer to the question is either 87 or 88. It seems certain that with 88 people $P(C) > 0.5$.