18.05 Problem Set 1, Spring 2014 Solutions

Problem 1. (10 pts.)
answer: (reasons below)
$P(\text{two-pair}) = .047539$, $P(\text{three-of-a-kind}) = 0.021128$, 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.)

We first choose two of thirteen ranks for the two pairs. $\binom{13}{2}$
For the pair of lower rank, we choose two of four suits. $\binom{4}{2}$
For the pair of higher rank, we choose two of four suits. $\binom{4}{2}$
For the singleton card, we choose one of eleven remaining ranks: $\binom{11}{1}$
For the singleton card, we choose one of four suits. $\binom{4}{1}$.

$$\frac{\binom{13}{2}\binom{4}{2}^2\binom{11}{1}\binom{4}{1}}{\binom{52}{5}} \approx .0475$$

Three-of-a-kind:
We choose one of thirteen ranks for the triple. $\binom{13}{1}$
We choose three of four suits for the triple. $\binom{4}{3}$
For the other two cards, we choose two of twelve remaining ranks. $\binom{12}{2}$
For the singleton of higher rank, we choose one of four suits: $\binom{4}{1}$
For the singleton of lower rank, we choose one of four suits: $\binom{4}{1}$

$$\frac{\binom{13}{1}\binom{4}{3}\binom{12}{2}\binom{4}{1}^2}{\binom{52}{5}} \approx .0211$$

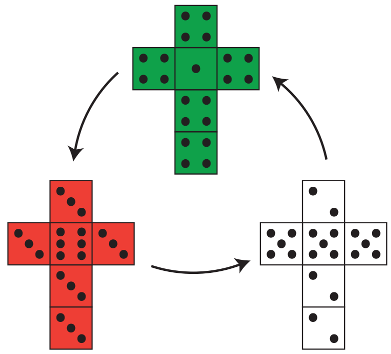
So two-pair is more than twice as likely as three-of-a-kind.

You can calculate the probability of other poker hands using a similar strategy. The full list is here: http://en.wikipedia.org/wiki/Poker_probability

Problem 2. (10 pts.)

answer: (reasons below)
(a) P(white beats green) = 7/12
P(green beats red) = 25/36
From class P(red beats white) = 7/12.

No you can't line them up since R beats W beats G beats R. 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.

	Red die		White die		Green 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		Green	
		2	5	1	4
Red	3	5/12	5/12	5/36	25/36
	6	1/12	1/12	1/36	5/36
Green	1	1/12	1/12		
	4	5/12	5/12		

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

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

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

	2 Red 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	25/72	25/144
2 Red	9	10/144	10/72	10/144
	12	1/144	1/72	1/144

We see, $P(\text{red wins}) = 59/144$, so $P(\text{white wins}) = 85/144$.

Problem 3. (20 pts.)

answer: The sample space Ω 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.

(a) 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) 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,

an outcome ω is in A if and only if $b_k = b$ for some index k 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) It's easier to calculate $P(A^c)$. There are 364^n outcomes in A^c since there are 364 choices for each birthday. So

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

We can find the size of the group needed for $P(A) > .5$ by trial and error, plugging in different values of n. Or we can set $P(A) = .5$ and solve for n.

$$1 - \frac{364^n}{365^n} = .5 \quad \Rightarrow \quad \left(\frac{364}{365}\right)^n = .5 \quad \Rightarrow \quad n \cdot \ln(\frac{364}{365}) = \ln(.5) \quad \Rightarrow \quad n \approx 252.65$$

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) While 365/2 different birthdays would have a 50 percent chance of matching your birthday, 365/2 people probably don't all have different birthdays, so they have a less than 50 percent chance of matching.

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

```
# colMatches is an 18.05 function. It needs to be in the working directory
# See http://ocw.mit.edu/ans7870/18/18.05/s14/r-code/rstuff.html
source('colMatches.r') # Set up the parameters
ndays = 365
npeople = 20
ntrials = 10000
sizematch = 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,sizematch) prob_B = mean(w)
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	.4123
30	.7007
40	.8913, .8867, .8926, .8865
41	.9005, .9003, .9037, .9006, .903

We see that the estimated probability of B is consistently less than .9 for npeople = 40 and consistently greater than .9 for npeople = 41. Therefore: **Answer:** 41.

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

.9333, .9333, .9, .8333, .9333, .8, 1

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

(f) It's easier to calculate $P(B^c)$, the probability that all n birthdays are distinct. Then 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) To estimate $P(C)$ we used the same code as in problem 3e, except we set sizematch = 3. Here is the table of (estimated) probabilities we found.

npeople	$P(C)$ (multiple values means multiple runs of the code)
20	.0121
30	.0293
50	.1274
80	.4241
90	.5389
85	.4832
86	.4841
87	.502, .4989, .4909, .5035
88	.5081, .5115, .5149, .5071

The estimates for 87 are sometimes above and sometimes below .5. The exact answer is either 87 or 88. It seems certain that with 88 people $P(C) > .5$.

(h) As in Lucky Larry (on the slides for class 2) in repeated trials, matched outcomes are more probable when some outcomes are more probable than others. (Think of the extreme case where one outcome has probably 1.)