Statistics 101 Summer I 2011
Homework #1

Adam Kapelner, Instructor

Due 9AM, Wednesday, June 1, 2011

Instructions and Philosophy

Once again, the path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual "working out".

Reading is still *required*. For this homework set, please read sections 7.2, 7.3, 8.1–8.4 in Foster & Stine as well as the first six pages of Donald Gillies "Philosophical Theories of Probability" chapter 1.[1]

The problems below are color coded: green problems are considered *easy*; yellow problems are considered *intermediate*, red problems are considered *difficult*; and purple problems are for *extra credit*. The *easy* problems are intended to be "giveaways" if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

This homework is worth 100 points but the point distribution will not be determined until after the due date. Late homework will be penalized 10 points per day up to a maximum of three days. After three days, it will receive a zero because I will post the solutions.

15 points are given as a bonus if the homework is typed using LaTeX. I posted a LaTeXhow-to on webcafe. If you are handing in homework this way, comment out this section and all pictures.

Set Theory   Problems below are related to set theory. The sets we talk about in class are composed of elements that are events. Some of the problems below will be about abstract sets that are divorced from the sets used in probability.

Problem 1   Consider the sample space Ω where you flip a fair coin and roll a fair die.

(a) Draw this event space in a box similar to how we did in class and indicate $|\Omega|$.

---

[1] http://www.amazon.com/Philosophical-Theories-Probability-Issues-Science/dp/041518276X

(b) Are singleton sets of the events in Ω mutually exclusive? collectively exhaustive?

(c) Does it matter if the coin is flipped before the die, after the die, or simultaneously with the die? Explain.

(d) Consider the set T which represents all events where the coin was flipped tails and E which represents the set of events where the die rolled an even number. Draw a Venn diagram and list the elements of the sets $T \cap E$ and $E \backslash T$ and mark them on the diagram.

(e) Describe fully the set $2^{(E \cup T)^C}$ *i.e.* list all its elements.

Problem 2 In the game Euchre[2], 24 playing cards are used consisting of only aces, kings, queens, jacks, tens, and nines.

(a) Construct Ω_E, the event space of a Euchre deck by using set notation and operations on Ω, the event space of a full deck. Use the "..." notation used in class to specify your sets explicitly.

(b) Let B be the set of black cards, F the set of face cards, ♠ the set of spades. Find the set on the right hand side:

$$\{A\spadesuit, 9\heartsuit, K\diamondsuit\} \cup \varnothing \;\subseteq\; \left((B \cup F)^C \cup \spadesuit\right)^C \backslash \{10\spadesuit, 10\clubsuit, 10\heartsuit\} \cap \Omega$$

Now, evaluate whether the statement itself is true.

Problem 3 This question will get your feet wet in JMP. Load the file `cars_mpg_hw1.JMP`. The column "name" is the make and model of the car and the column "HWYMPG" is the gas mileage on the highway in miles per gallon.

(a) The "HWYMPG" column's data prints extra significant digits; fix this. Now, sort the dataset by gas mileage with the *highest* first. Now truncate our list of cars so you only have 40 total. Now sort alphabetically by make and model. Last, create a new column that lists the highway mileage in kilometers per liter and display to three significant digits. Print this table and attach it to your homework.

(b) Using the truncated list from part (a), let A be the set of all cars with less than 32 MPG or greater than or equal to 40 MPG. Find A^C and intersect it with the set of all Toyotas or all Hyundais. List these cars below.

Problem 4 In mathematics, we use convenient notation to represent intervals of real numbers. We represent all numbers $1 \le x \le 4$ by the bracket notation $[1, 4]$ the square bracket is for "closed" (including the endpoint) and the open parentheses are for "open" (not including the endpoint) so $1 < x < 4$ would be $(1, 4)$.

---

[2]see `http://en.wikipedia.org/wiki/Euchre` for more information

(a) Draw a number line for x and shade in the area that represents the set $[1,3] \cup [4,9]$

(b) Draw a number line for Z and share in the area that represents $|Z| \geq 2$. This Z notation we'll be using in a couple weeks, so I want to make sure you're familiar with this letter in our glorious alphabet.

(c) Draw on the number line the set $[0,1] \cap \left[0, \frac{1}{2}\right] \cap \left[0, \frac{1}{4}\right]$.

(d) Describe the set $\bigcup_{i=1}^{\infty} \left[0, \frac{1}{2^i}\right]$.

(e) Describe the set $\bigcap_{i=1}^{\infty} \left[0, \frac{1}{2^i}\right]$. Hint: imagine if the top bound was not ∞ but something like 3 and draw it out until you see a pattern.

Counting Problems below are related to counting. We will review the methods learned in class and expand our horizons.

Problem 5 We will roll different types of dice.

Let R be a standard 6-sided die, let S be an 8-sided die, let T be a 12-sided die, and let U be a 20-sided die. What is the sample size of Ω for the experiment where we...

(a) roll R 3 times?

(b) roll R then S then T then U?

(c) roll R 53 times, then wait 53 days, then roll S 32 times and climb 32 meters, then roll T 47 times and drive 47 miles, then roll U 87 times and watch 87 movies.

Problem 6 Examine the following words and tell me how many *permutations* there are of the letters. We do not care about keeping track of the individual common letters. For example, in the word *dad*, there are two $d's$ and we want to treat the permutation $d_1 d_2 a$ the *same* as $d_2 d_1 a$.

(a) town

(b) mississippi

(c) supercalifragilisticexpialidocious[3]

Problem 7 Below is a standard chessboard.[4] In the following problems, pieces of the same type are considered interchangeable (*i.e.* all white pawns are the same, all black rooks are the same, etc) but *not* across color (a white pawn is different from a black pawn).

(a) How many ways are there to place the white king on a white square?

(b) How many ways are there to set up the pieces in the back ranks of both white and black *i.e.* arrange the two rooks, two knights, two bishops, king and queen on the first row of 8 squares.[5]

(c) How many ways are there to arrange the pieces on the board?

Problem 8 We have 4 blue marbles, 4 green marbles, 2 orange marbles, and 2 red marbles.

---

[3]see http://en.wikipedia.org/wiki/Longest_word_in_English

[4]If you are not familiar with the pieces in chess, please read http://en.wikipedia.org/wiki/Chess

[5]This is called "Fischer Random Chess" after the famous grandmaster Bobby Fischer who proposed the idea for a more fun game.

For all these questions, if you are using "choose notation", please write your choose notation, then write the formula using factorials, then write the actual number after you compute it.

(a) Viewing all the marbles as *unique*, how many ways is there to order the marbles?

(b) Viewing all marbles of the same color as *interchangeable*, how many ways is there to order the marbles?

(c) If I pick 4 marbles at random from the collection, what is the probability of getting all four marbles the same color?

(d) If I pick 4 marbles at random from the collection, what is the probability of getting two greens and one blue and one orange?

(e) If I pick 4 marbles at random from the collection, how many ways are there to get two-of-a-kind *i.e.* two marbles of one color and two marbles of a different color.

Probability   This section will cover probability foundations (addition rule, multiplication rule, complements) as well as conditional probability.

Problem 9   A businessman has a meeting in one of the three tallest buildings in downtown Philadelphia.

From left to right we have the Comcast Center at 1701 John F. Kennedy Blvd (58 floors), One Liberty Place at 1650 Market Street (61 floors), and Two Liberty Place at 1601 Chestnut Street (58 floors). Consider the case where he enters a random building and goes to a random floor

(a) Draw a probability tree (see p.183 of Stine & Foster) of this random event. Use "..." notation so your trees don't take up the whole page.

(b) Are the building selection and floor selection *independent*?

(c) What is the probability of the businessman winding up on floor 23 of Two Liberty Place?

(d) What is the probability of the businessman winding up on floor 23 of any building?

(e) What is the probability of the businessman winding up on floor 60 of any building?

(f) If the businessman is on floor 60, what is the probability he is in One Liberty Place?

(g) What are the odds he winds up on any floor between 1 through 58?

(h) If the businessman is on floor 5, what is the probability he is in One Liberty Place?

(i) Let's say the businessman does this straight for four weeks, five days per week. What is the *odds* he winds up on floor 32 every business trip?

Problem 10 This will get your feet wet using R.[6]

(a) To calculate combinations, use the `choose(n,k)` function. Calculate the number of five-card hands from a standard deck by copying the following code into R and then pressing enter:

```
choose(52, 5)
```

Please write down the answer.

(b) Verify the probability in class of a "full house" by copying the following code into R and then pressing enter:

```
choose(13, 1) * choose(4, 3) * choose(12, 1) * choose(4, 2) / choose(52, 5)
```

Write down the answer as a *percentage*.

(c) We are going to do a little experiment to explore the definition of probability as a limiting frequency. We will be looking at the context of flipping a coin and getting heads. Remember the definition was:

$$\mathbb{P}\left(H\right) = \lim_{n \to \infty} \frac{\text{count}(\text{H in } n \text{ flips})}{n}$$

Run the following code by copying and pasting:

---

[6]Please download it from `http://cran.r-project.org/` and then double-click to open an R console.

```
N = 30000
sims = sample(0:1, N, replace = T)
freqs_by_n = array(NA, N)
for (n in 1 : N){
  freqs_by_n[n] = sum(sims[1:n]) / n
}
plot(10:N,
  freqs_by_n[10:N],
  xlim = c(10, N),
  ylim = c(0.40, 0.60),
  pch = ".",
  xlab = "number of samples",
  ylab = "frequency of heads",
  main = "P(H) as a limiting frequency: 30,000 samples")
abline(h = 0.5, col = "blue")
freqs_by_n[N]
#last line placeholder
```

The console should have popped up a plot.[7] Print this out and attach it to your homework.

From the title of the plot and the x and y axes, tell a story about what is going on here.

What is the limiting frequency of heads after 30,000 coin flips to 3 decimals? (that is the number that appears in the console directly after "> freqs_by_n[N]")

Problem 11 As part of a health study, 987 patients were sampled and 296 were found to be obese (denote this set L for "large"), 82 were found to be diabetic (call this set D) and 49 were found to be *both* obese and diabetic.
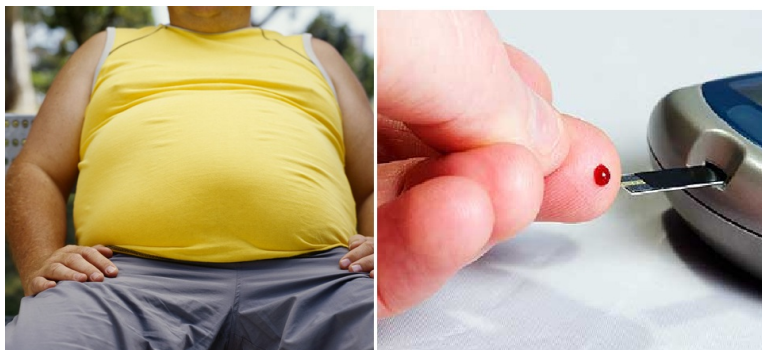
(a) Draw Ω, $L \subset \Omega$, $D \subset \Omega$ as a Venn diagram *to scale* as best as you can.

_____

[7]This is a *real* statistical simulation. Each time you run this code it will be different. You can compare plots with your friends but take note that they will not look exactly the same.

(b) Draw a table of counts similar to the one on p.175 of Stine & Foster.

(c) Are being obese and being diabetic dependent or independent and justify your answer giving insight into what you think is happening from a medical point of view.

(d) Draw a table of frequencies similar to the one on p.175 of Stine & Foster.

(e) Estimate $\mathbb{P}(L)$, $\mathbb{P}(D)$, $\mathbb{P}(L, D)$ by assuming 987 is a long-run frequency.

(f) Find $\mathbb{P}(D|L)$ and interpret the question and answer in your own words.

(g) Find $\mathbb{P}(L|D)$ and interpret the question and answer in your own words.

(h) Find $\mathbb{P}(D|L^C)$ and interpret the question and answer in your own words.

(i) Draw a probability tree of this space similar to the one on p.183 of Stine & Foster.

Problem 12 Philosophy

(a) On p.1 of Gillies book, explain the "four principle current interpretations" in your own words.

8