Axioms and assumptions

18.600 Problem Set 2

Welcome to your second 18.600 problem set! We will continue to explore some combinatorics along with probability and the axioms of probability.

Before we get to work, let's indulge in a bit of reflection. When we say "The probability that A will happen is p" where does p come from? Sometimes the evidence convinces pretty much everyone that A will or will not happen. Informally, the probability that a predicted lunar eclipse will happen on schedule is pretty much 1, and the probability that Mars and Jupiter will collide this month is pretty much 0. In other simple situations (die rolls, coin tosses, etc.) experience may lead us to agree on probabilities that aren't 0 or 1. The assumption that all outcomes are equally likely (for random permutations or die rolls or coin tosses) is sometimes a natural starting point. This assumption is implicitly made in a few of the problems here.

In more complicated real world settings, one can sometimes define the *risk neutral* probability, a probability measure derived from the market prices of contracts whose values depend on future events. If we want to know the risk neutral probability that a given candidate will win an election, or that an athletic team will win a game, we can look at betting markets. (Check out electionbettingodds.com, predictit.com, oddschecker.com, betfair.com, and similar websites.) As we will see later in the course, if we want to know the risk neutral probability that the price of a share of Apple stock will exceed some value by the end of the year, we can work this out by looking at current prices of *derivatives* (contracts whose future value depends on future share prices). The total amount of money at stake in derivative markets is estimated at over a quadrillion dollars per year (try googling derivatives quadrillion).

Some argue that betting markets set up perverse incentives. If I buy a contract that gives me $\$500,000$ if my house burns down, that's useful insurance. But if I buy a contract that gives me $\$500,000$ if *your* house burns down, that gives me an unhealthy incentive to burn your house down. People similarly worry about a world in which hedge funds can bet that a company will collapse and then actively cause it to collapse. Rules are required to prevent such things, but foolproof (and evil-genius-proof) rules are hard to design and enforce.

On the other hand, one might argue that the absence of betting markets is part of the reason that some questions in politics and law are divisive. It is hard to place a bet on the proposition that "my candidate would do more to advance long term happiness and prosperity than yours" or "my client is innocent," so there is no market mechanism for producing a commonly accepted probability. Different groups can *claim* to have different probability estimates, the expression of which may advance their own agendas, but without a market we cannot tell which parties would actually be *willing to bet money* at the corresponding rates. Some studies claim that people answering questions about the economy are both more accurate and less partisan when they are paid (even a very small amount) for correct answers. Maybe there is something to be said for having money on the line.

A. FROM ROSS 8TH EDITION CHAPTER TWO:

1. **Problem 25:** A pair of dice is rolled until a sum of either 5 or 7 appears. Find the probability that a 5 occurs first. *Hint.* Let E_n denote the event that a 5 occurs on the nth roll and no 5 or 7 occurs on the first $(n-1)$ rolls. Compute $P(E_n)$ and argue that $\sum_{n=1}^{\infty} P(E_n)$ is the desired probability.
2. **Theoretical Exercise 10:** Prove that
$P(E \cup F \cup G) = P(E) + P(F) + P(G) - P(E^cFG) - P(EF^cG) - P(EFG^c) - 2P(EFG)$.
3. **Theoretical Exercise 20:** Consider an experiment whose sample space consists of a countably infinite number of points. Show that not all points can be equally likely. Can all points have a positive probability of occurring?

B. On Sep. 12, 2019, I looked up Democratic presidential nomination contracts on two trading platforms: predictit and betfair. According the table below, one could purchase a predictit YES contract on Elizabeth Warren for 35 cents. This is worth \$1 if Warren wins the nomination and \$0 otherwise. Similarly, one could purchase a predictit NO contract on Elizabeth Warren for 66 cents. This is worth \$1 if Warren is *not* elected and \$0 otherwise. (When you purchase a NO contract for 66 cents you are technically *selling* a YES contract for 34 cents to somebody else offering to buy it for that price. The gap between the 34 offer-to-buy and the 35 offer-to-sell is called the *bid-ask spread.* Because predictit only trades in integers, the spread is typically 1 cent — which is why YES and NO prices sum to 101 instead of 100.)

	predictit YES	predictit NO	betfair YES	betfair NO
Elizabeth Warren	35	66	34.5	66.2
Joe Biden	27	74	21.7	79.2
Bernie Sanders	17	84	13.5	87.2
Andrew Yang	12	89	6.1	94.6
Kamala Harris	11	90	10.9	89.6
Pete Buttigieg	9	92	5.0	95.8
Hillary Clinton	5	96	3.6	96.7
Cory Booker	4	97	2.5	98.3
Beto O'Rourke	3	98	0.8	99.6
Julian Castro	2	99	1.0	99.1
Tulsi Gabbard	3	98	1.0	99.3
Tom Steyer	3	98	0.8	99.9
Amy Klobuchar	2	99	0.7	99.5

Because the betting probabilities are not the same on the two sites, there may be opportunities for *arbitrage*, i.e., opportunities to make money without taking risk. (For now, let us ignore taxes, fees, interest and the risk associated with either of these companies going bankrupt before the election; of course these things matter in practice.) If I buy Biden NO on predictit for 74 and Biden YES on betfair for 21.7 then I have only spent 95.7. But I am guaranteed to win 100. If I purchase all 13 predictit NO contracts then the price comes to \$11.80. But because at least 12 of these candidates are guaranteed to lose, I will be guaranteed to win at least \$12 before fees (and about \$11.88 after fees). Note that per predictit rules, I only have to leave enough money on the site to cover the amount I will owe in the worst case scenario; this means that I can take my 8 cents of profit now *without* having to leave any money on the platform for the duration between now and the election. So: 8 free cents.

1. Using the chart above, find three other arbitrage opportunities, i.e., ways to make a risk free profit (again ignoring the caveats mentioned above: fees, tax, interest, etc.)
2. Per the efficient market hypothesis, arbitrage opportunities should not exist. Give a short speculative explanation for why there seem to be arbitrage opportunities here. (This will not be graded except to check that you thought about it.)

REMARK: Predictit traders are only allowed to spend \$850 on each candidate. (This was part of the deal making predictit legal in the US, unlike betfair.) So one can buy 42500 shares of Klobuchar YES (2 cents each) but only 858 shares of Klobuchar NO (99 cents each). Some people think these constraints cause unlikely YES contracts to be overpriced. In theory, arbitrageurs should correct this overpricing by buying lots of NO contracts (thus bidding down the YES prices); but if each potential arbitrageur can only make a *little* "free money" this way (due to spending caps), it may just not be worth the effort.

REMARK: Suppose that you think the "true" value of Warren YES is 34.5 but have a personal reason for wanting to own a Warren YES share. Then there are two things you can do. One, you can go straight to the site and buy a share for 35 (basically overpaying by half a cent). Or two, you can put up an "offer" to buy for 34 and wait until somebody is willing to take the other side of the bet (by buying NO at 66). If you take the second approach, you might get a better price. But you will have to wait in line behind everyone else offering to buy at 34. And there is some risk that you never get to the front of the line (perhaps the price of Warren YES shares goes up before you manage to buy at 34) — and the scenarios in which you fail to get to the front of the line might be precisely the scenarios in which you would most like to own Warren YES. Sometimes on predictit the lines can be long (tens or even hundreds of thousands of shares) and because trades can only take place at integer values, you cannot jump to the front of the line by offering 34.1. For example, last I checked there are over 350,000 offers to *sell* Mark Zuckerberg YES for 1 cent. (Zuckerberg is not running.) Some people making those offers would be willing to sell for half a cent or a tenth of a cent. But because of the strict integer rule, they can't make those offers. They have to wait in line.

C. Suppose that there are $M \geq 1$ job candidates and $N \geq 1$ companies with job openings. Each candidate (independently, uniformly at random) develops a serious interest in one of the

N companies and each company (independently, uniformly at random) develops a serious interest in one of the M candidates. What is the probability that there is at least one company-candidate pair that are seriously interested in each other? (Hint: let E_j be the event that the jth applicant's interest is requited. Use inclusion-exclusion on these events. You can write the probability as a sum. Don't worry about simplifying further.) Later in this course (after we discuss *additivity of expectation*) we will find that the *expected* (a.k.a. *average*) number of candidates with requited interest is $\sum_{j=1}^{M} P(E_j)$. Compute this quantity as a function of M and N, and write a sentence about whether you consider the answer surprising.

D. Alice and Bob are playing a game of tennis and have reached the game state called "deuce." From here the players keep playing points until one player's point-win total exceeds the other player's total by 2, at which point the player ahead by 2 is declared winner of the game. Suppose that Alice wins each point with probability p (independently of all previous points) and Bob wins each point with probability $q = (1 - p)$ (independently of all previous points). Find the probability that Alice wins the game, as a function of p. (Hint: consider what happens over the course of the next *two* points. Either Alice wins both and the game is over, or Bob wins both and the game is over, or each player wins a point and the players are back where they started. Compute the probabilities of these three outcomes. Then apply the ideas from the first problem on this problem set.) Based on your answer, do you agree or disagree with the following statement? *If Alice is k times as likely as Bob to win a point, then Alice is k^2 times as likely as Bob to win the game if the current score is deuce.*

E. The online comic strip xkcd.com has a "random" button one can click to choose one of the previous $n \approx 2200$ strips. Assume there are exactly 2200 numbered strips and that clicking the "random" button yields the kth strip, where k is chosen uniformly from $\{1, 2, \ldots, n\}$. If one observes m strips this way, what is the probability that one sees at least one strip more than once? (This is a variant of the birthday problem.) Give rough numerical values for $m = 36$ and $m = 56$ and $m = 78$. (Hint: try going to wolframalpha.com and entering something like Prod[(1-k/2200), {k,0,35}].) Based on your answer, do you agree or disagree with the following statement? *The number of clicks required before you see the same strip twice is a random quantity whose median is about* 56*, and which lies between* 36 *and* 78 *about half the time*

F. A "gender reveal" party is held to announce the gender of an expected newborn. 15 cups are filled in advance with colored beads and covered: if the baby is a boy, 8 are filled with blue beads and 7 with pink. If the baby is a girl, 7 are filled with blue and 8 with pink. When the audience arrives the cups are knocked over (revealing bead colors) in a uniformly random order until the audience has seen 8 cups of the same color (and thereby knows the gender). Let N be the number of cups turned over before the gender is known. Compute the probability that $N = k$ for $k \in \{8, 9, 10, \ldots, 15\}$. Is the probability that $N = 15$ (and one has to wait to the very end) more or less than 1/2? (This problem was communicated to me by a friend who found $N = 15$ in a real life implementation and wanted to know how surprising that was.)

G. The following is a popular and rather instructive puzzle. A standard deck of 52 cards (26 red and 26 black) is shuffled so that all orderings are equally likely. We then play the following game: I place the deck face down and begin turning over the cards from the top of the deck one at a time so that you can see them. At some point (before I have turned over all 52 cards) you say "now." At this point I turn over the next card and if the card is red, you receive one dollar; otherwise you receive nothing. You would like to design a strategy to maximize the probability that you will receive the dollar. How should you decide when to say "now"?

Your first observation is that a good time to say "now" is when you know a high fraction of the cards remaining in the deck are red. On the other hand, if you wait for this fraction to increase, there is a chance you'll see more red cards while you wait, so that the fraction *decreases*. What's the right way to balance these concerns, i.e., what is the *optimal* strategy for deciding when to say "now"? The next page has a hint but don't look until you have to.

HINT: Imagine a variant of the game in which, after you say "now," I turn over the *bottom* card on the deck. Observe that in this variant, it makes no difference when you say "now" (since you're going to see the same bottom card of the deck regardless). Now try to argue that your probability of winning with a given strategy in the modified game is the same as your probability of winning with that strategy in the original game. Conclude that in the original game, it also makes no difference which strategy you choose. Your odds of winning the dollar are .5 regardless.

REMARK: The fraction of red cards in the deck turns out to be something called a *martingale* and the fact that it makes no difference when you bet can be derived from something called the *optional stopping theorem*. We'll have more on this at the end of the course. But I like this puzzle because it's something you can appreciate right now. This is a strategy game that Warren Buffett and a chimpanzee would play equally well. Unlike the game shown here `https://www.youtube.com/watch?v=JkNV0rSndJ0` and playable here `http://www.softschools.com/games/memory_games/ayumus_game/`