CS 70 Discrete Mathematics and Probability Theory
Fall 2013 Vazirani Week 8 Discussion

Sample Spaces, Basic Probability and Conditional Probabilities

Suppose that one of your friends is starting a Combinatorics Club, and n people have signed up to join. The club is still in its formative stages, and you need to hold elections in order to choose the club's leadership. Your friend decided that the club should be governed as follows: there should be one club president, plus k cabinet members who are all equal in power, and they assist the president.

1. **At first, your friend wants to nominate $k+1$ people ahead of time, and one of these people will be randomly selected to be president; the rest will be cabinet members. Let the nominees be $p_1, p_2, \ldots, p_{k+1}$.**

   (a) **What is the *sample space* for the leadership groups?** There are $k+1$ sample points in the sample space. Sample point i represents the outcome of p_i being chosen as president and $p_1, \ldots, p_{i-1}, p_{i+1}, \ldots, p_{k+1}$ as cabinet members.

   (b) **What is one *outcome* in this sample space?** See above.

   (c) **Give one example of an *event* in this sample space.** A possible event would be all leadership groups with either p_1 or p_2 as president.

   (d) **Assuming you are one of the nominees, what is the probability that you will be part of the leadership (president or cabinet member)?** Notice that in all sample points, you are either president (in 1 sample point) or cabinet member (in k sample points). The probability that you are either president or cabinet member is therefore 1.

   (e) **Assuming you are one of the nominees, what is the probability that you will be president?** This occurs in 1 out of the $k+1$ sample points - so the probability is $\frac{1}{k+1}$.

   (f) **Assuming you are one of the nominees, what is the probability that you will be a cabinet member?** This occurs in k out of the $k+1$ sample points - so the probability is $\frac{k}{k+1}$.

   (g) **Now, assuming instead that the nominees were initially chosen uniformly at random, what is the probability you will be president?** Let's think about this in terms of the sample space. We have $\binom{n}{k+1}$ copies of the leadership group sample space described above - each of these copies corresponds to a different choice of the $k+1$ nominees. That means there are $\binom{n}{k+1} \times (k+1)$ sample points total. There are $\binom{n-1}{k}$ copies in which you are one of the nominees (this is the number of ways of choosing k of the remaining $n-1$ members to be the other k nominees). In each of these copies, there are $k+1$ sample points, as described above, and in one of those sample points you are president. So the number of sample points in which you are president is $\binom{n-1}{k}$. The probability that you are president is therefore $\frac{\binom{n-1}{k}}{\binom{n}{k+1} \times (k+1)} = \frac{1}{n}$.

2. **Your friend changes her mind, and decides that it is best to allow each of the n people signed up to run for president, and to allow each of the candidates to choose a vice president.**

   Alice, Bob, Chuck, and Delilah all sign up as candidates, and each has a list of people they equally like. They will select their vice president randomly from this list.

Candidate	Vice President Options
Alice	Emily, Fred, Gary, Hilda
Bob	Alice, Emily, Gary, Jordan, You
Chuck	Alice, Bob, You
Delilah	Chuck, Gary, Iggy, Jordan

Notice that these lists overlap (for example, you can be either Bob or Chuck's VP).

(a) **What is the *sample space* for the leadership groups?** Each sample point consists of a possible president and vice president pair. There are 16 sample points- 4 with Alice as president, 5 with Bob as president, 3 with Chuck, and 4 with Delilah.

(b) **What is one *outcome* in this sample space?** An outcome would be a possible president and vice president pair, like Alice as president and Emily as vice president.

(c) **Give one example of an *event* in this sample space.** One possible event would be the event that Emily is vice president. This would consist of two sample points - one with Bob as president and Emily as vice, and one with Alice as president and Emily as vice.

(d) **Assuming the winning candidate is chosen uniformly at random, what is the probability that Gary will be part of the leadership (president or VP)?** How many sample points include Gary as part of the leadership? Three sample points - when Alice, Bob or Delilah is president. However, all three of these sample points are not assigned the same probability. The four sample points with Alice as president have probability $\frac{1}{16}$ each - Alice is chosen as president with probability $\frac{1}{4}$, and each of her vice presidents is then chosen with probability $\frac{1}{4}$. The five sample points with Bob as president have probability $\frac{1}{20}$ each. The three sample points with Delilah as president have probability $\frac{1}{16}$ each. Therefore, this is not a uniform probability space. The probability that Gary will be part of the leadership is then $\frac{1}{16}+\frac{1}{20}+\frac{1}{16}=\frac{7}{40}$.

(e) **Assuming Emily becomes vice president, what is the probability that Gary is part of club leadership?** Now we only consider the sample points in which Emily is vice president. Note that Gary can only be vice president - he is not running for president. Emily and Gary cannot be vice president at the same time, so the probability that Gary is part of club leadership is 0.

(f) **Assuming that Alice has a %30 chance of winning, Bob has a %40 chance of winning, Chuck has a % 20 chance of winning and Delilah has a %10 chance of winning, what is the probability that you are part of the club leadership?** You are only vice president in two sample points - either Bob or Chuck must win the election and then they must select you as vice president. You are not running for president. The probability of the sample point in which Bob is president and you are vice is $\frac{4}{10}\times\frac{1}{5}$ - Bob wins with probability $\frac{4}{10}$, and selects you with probability $\frac{1}{5}$. The probability assigned to the sample point in which Chuck is president and you are vice is $\frac{2}{10}\times\frac{1}{3}$. The probability that you are part of the club leadership is then $\frac{11}{75}$.

(g) **Suppose that due to a scandal, Delilah drops out of the election. Conditioning on the fact that Delilah does not win, and given that Alice originally had a %30 chance of winning, Bob had a %40 chance of winning, and Chuck had a %20 chance of winning, what is the probability that Alice is part of the club leadership?** We are now consider a smaller sample space - the sample space consisting of only the points in which Delilah is not president. Instead of 16 sample points, there are now only 12. If we consider the probabilities assigned to each of the remaining sample points, the total probability is now only $\frac{9}{10}$ instead of 1 (because the sum of the probabilities assigned to the sample points with Delilah as president is $\frac{1}{10}$).

There are 3 possible situations in which Alice can be part of club leadership. She could be chosen as president, which occurs with probability $\frac{3}{10}$. Bob could be chosen as president and then choose her as vice president. This occurs with probability $\frac{4}{10} \times \frac{1}{5}$. Finally, Chuck could be chosen as president and then choose her as vice president. This occurs with probability $\frac{2}{10} \times \frac{1}{3}$. If we sum the probabilities of all three of these situations, we obtain $\frac{67}{150}$. However, we must divide this number by $\frac{9}{10}$ rather than 1, since we have restricted the sample space. The probability that Alice is part of club leadership is then $\frac{67}{135}$.

3. **Your friend decides that Alice, Bob, Chuck and Delilah are not really fit to lead Combinatorics Club.**

   In a stroke of luck, Delilah drops out of the race due to a scandal, Chuck drops out because he decides that being president would be too much on top of his courseload, and Bob drops out to spend more time with his family. So it seems like Alice will win by default.

   To avoid this, your friend enters you into the election as a candidate, and she will rig the election so that you are more likely to win. When the votes come in, starting from day 0, every day she will sample a vote by drawing it uniformly at random from a hat. If the vote is for you she will declare you the victor. If the vote is for Alice, then if it is an even-numbered day she will return it to the hat, but if it is an odd-numbered day she will declare that Alice won.

   (a) **Assuming there were an equal number of votes for you and for Alice, what is the probability that you win the election?** On day 0, you win with probability $\frac{1}{2}$, and with probability $\frac{1}{2}$ your friend draws a vote for Alice, which means the vote is returned to the hat. In this case, the voting continues to day 1. On day 1, you again win with probability $\frac{1}{2}$. With probability $\frac{1}{2}$, your friend draws a vote for Alice - but this time, she declares Alice as the winner, since day 1 is an odd numbered day.

   To determine the probability that you win the election, let's consider the sample space. Let each sample point consist of a pair of votes - the first element in the pair represents the vote drawn on day 0, and the second element in the pair represents the vote drawn on day 1. The first element can either be Alice or you. The second element can be Alice, you, or $\emptyset$ (which means the voting stopped on day 0). The sample points are: $(you, \emptyset), (Alice, you), (Alice, Alice)$. The first sample point occurs with probability $\frac{1}{2}$ - this is the probability a vote for you is drawn on day 0. The second sample point occurs with probability $\frac{1}{2} \times \frac{1}{2}$ - this is the probability that a vote for Alice is drawn on day 0 and a vote for you is drawn on day 1. The third sample point occurs with probability $\frac{1}{2} \times \frac{1}{2}$. It is the probability that votes for Alice are drawn on both day 0 and day 1. The probability that you win is the probability that either the first or second sample point occurs - this is $\frac{1}{2} + \frac{1}{4} = \frac{3}{4}$.

   (b) **Assuming that Alice had %70 of the vote, what is the probability that you win the election?** The sample space is the same as above, but the sample points are assigned different probabilities. The first sample point is assigned probability $\frac{3}{10}$ (the probability that a vote for you is drawn on the first day). The second sample point is assigned probability $\frac{7}{10} \times \frac{3}{10}$. The third sample point is assigned probability $\frac{7}{10} \times \frac{7}{10}$. The probability that you win is still the sum of the probabilities of the first two sample points - so $\frac{3}{10} + \frac{21}{100} = \frac{51}{100}$.

   (c) **Extra: What is the minimum proportion of the vote that Alice needs in order to win the election with probability at least $\frac{1}{2}$?** Let x be Alice's proportion of the vote. The sample points are the same as those defined above. Then the first sample point is assigned probability $1-x$, the second sample point is assigned probability $x(1-x)$, and the third sample point is assigned probability x^2. You win with probability $1-x+x(1-x) = 1-x^2$, so Alice wins with $1-(1-x^2) = x^2$. We would like $x^2 \geq \frac{1}{2}$, so x must be at least $\frac{1}{\sqrt{2}} \approx .707$.

4. **Challenge: Recall the Monty Hall Problem from lecture: A contestant is participating in a game show, in which there are 3 doors. Behind two of the doors are goats, behind one of the doors is a prize; if the contestant chooses the door with the prize behind it, they can keep the prize.**

   When the contestant selects a door, the game show host opens one of the other two doors to reveal a goat. Then, the contestant has the option of changing doors.

   What is the probability that the contestant gets the prize if they stay at the current door? What is the probability that the contestant gets the prize if they switch to the other door?

   This can be done using conditional probabilities, which you will see in note 10. However, we will now work through this using sample spaces.

   What is the sample space here? We can describe the outcome of the game (up to the point where the contestant makes his final decision) using a triple of the form (i, j, k), where $i, j, k \in \{1, 2, 3\}$. The values i, j, k respectively specify the location of the prize, the initial door chosen by the contestant, and the door opened by Carol. Note that some triples are not possible: e.g., $(1, 2, 1)$ is not, because Carol never opens the prize door. Thinking of the sample space as a tree structure, in which first i is chosen, then j, and finally k (depending on i and j), we see that there are exactly 12 sample points.

   Assigning probabilities to the sample points here requires pinning down some assumptions:

   - The prize is equally likely to be behind any of the three doors.
   - Initially, the contestant is equally likely to pick any of the three doors.
   - If the contestant happens to pick the prize door (so there are two possible doors for Carol to open), Carol is equally likely to pick either one. (Actually our calculation will have the same result no matter how Carol picks the door.)

   From this, we can assign a probability to every sample point. For example, the point $(1, 2, 3)$ corresponds to the prize being placed behind door 1 (with probability $\frac{1}{3}$), the contestant picking door 2 (with probability $\frac{1}{3}$), and Carol opening door 3 (with probability 1, because she has no choice). So

   $$\Pr[(1,2,3)] = \frac{1}{3} \times \frac{1}{3} \times 1 = \frac{1}{9}.$$

   [Note: Again we are multiplying probabilities here, without proper justification!] Note that there are six outcomes of this type, characterized by having $i \neq j$ (and hence k must be different from both). On the other hand, we have

   $$\Pr[(1,1,2)] = \frac{1}{3} \times \frac{1}{3} \times \frac{1}{2} = \frac{1}{18}.$$

   And there are six outcomes of this type, having $i = j$. These are the only possible outcomes, so we have completely defined our probability space. Just to check our arithmetic, we note that the sum of the probabilities of all outcomes is $(6 \times \frac{1}{9}) + (6 \times \frac{1}{18}) = 1$.

   Let's return to the Monty Hall problem. Recall that we want to investigate the relative merits of the "sticking" strategy and the "switching" strategy. Let's suppose the contestant decides to switch doors. The event A we are interested in is the event that the contestant wins. Which sample points (i, j, k) are in A? Well, since the contestant is switching doors, his initial choice j cannot be equal to the prize door, which is i. And all outcomes of this type correspond to a win for the contestant, because Carol must open the second non-prize door, leaving the contestant to switch to the prize door. So A consists of all outcomes of the first type in our earlier analysis; recall that there are six of these, each with probability $\frac{1}{9}$. So $\Pr[A] = \frac{6}{9} = \frac{2}{3}$. That is, using the switching strategy, the contestant wins with probability $\frac{2}{3}$! It should be intuitively clear (and easy to check formally — try it!) that under the sticking strategy his probability of winning is $\frac{1}{3}$. (In this case, he is really just picking a single random door.) So by switching, the contestant actually improves his odds by a huge amount!