DSC 40A - Probability Roadmap

A hopefully helpful guide to solving probability questions

1. What is the sample space, S (all possible outcomes)? Consider using:

- individual objects
- sequences of objects
- sets of objects
- sequences of positions
- sets of positions

Example: Suppose that a standard deck of 52 cards is shuffled in a random order. What is the probability that both red queens are adjacent?

Try to do this problem as many ways as you can, using a different sample space each time.

We could use:

- S = sequences of 52 cards
- S = ordered pairs (sequences) of positions where the red queens are
- S = sets of two positions where the red queens are

Note: sets of 52 cards doesn't make sense as a sample space, because we always have the same set of 52 cards in a shuffled deck. The order is what changes, and sets don't capture that.

Each way of setting up the sample space may lead to a different way of obtaining the answer. Some may be easier than others, like the "Easy Way" of sampling without replacement discussed in Lecture 23. But all should result in the same numerical answer.

- If S = sequences of 52 cards, then the probability of both red queens being adjacent is
$$\frac{51 * 2 * 50!}{52!} = \frac{1}{26}$$
.
- If S = ordered pairs (sequences) of positions where the red queens are, then the probability of both red queens being adjacent is
$$\frac{51 * 2}{52 * 51} = \frac{1}{26}$$
.
- If S = sets of two positions where the red queens are, then the probability of both red queens being adjacent is
$$\frac{51}{C(52, 2)} = \frac{1}{26}$$
.

For more details on the various solutions to this problem, refer to the materials from Synchronous Class 7.

2. Are all of the outcomes in S equally likely, or are some more likely than others?

If all equally likely (most common), can use

$$P(\text{favorable outcome}) = \frac{\text{number of favorable outcomes in } S}{\text{total number of outcomes in } S}.$$

Then our probability question turns into two counting questions.

If not all equally likely, add up probabilities associated with each of the favorable outcomes, using

$$P(E) = \sum_{s \text{ in } E} P(s).$$

Example: Consider the sample space $S = \{t, u, v, w, x, y, z\}$ with associated probabilities given in the table below.

outcome	t	u	v	w	x	y	z
probability	$\frac{5}{21}$	$\frac{2}{21}$	$\frac{1}{21}$	$\frac{4}{21}$	$\frac{2}{21}$	$\frac{4}{21}$	$\frac{3}{21}$

Let $A = \{t, u, v, w\}$ and $B = \{v, w, x, y\}$. Find $P(A|B)$.

Use the formula for conditional probability:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

For the numerator, $A \cap B = \{v, w\}$, so

$$P(A \cap B) = P(v) + P(w) = \frac{1}{21} + \frac{4}{21} = \frac{5}{21}.$$

For the denominator,

$$P(B) = P(v) + P(w) + P(x) + P(y) = \frac{1}{21} + \frac{4}{21} + \frac{2}{21} + \frac{4}{21} = \frac{11}{21}.$$

Putting this together,

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{5}{21}}{\frac{11}{21}} = \frac{5}{11}.$$

3. What are the favorable outcomes in S? Try to write down one favorable outcome, then think about what was forced and what you were free to select in generating this example. This can help you count the number of favorable outcomes.

Example: If favorable outcomes are sequences of 52 cards with both red queens adjacent, list one favorable outcome and count the number of favorable outcomes.

One favorable outcome is:

$$10\heartsuit, 3\clubsuit, 7\spadesuit, Q\diamondsuit, Q\heartsuit, A\diamondsuit, \ldots, 3\spadesuit$$

In generating an example like this, I get to choose:

- where the first red queen goes $\longleftarrow$ 51 options
- whether the first red queen is Q♢ or Q♡ $\longleftarrow$ 2 options
- what card to place in each of the remaining 50 positions $\longleftarrow$ 50! options

I don't get to choose:

- what card goes after the first red Q (it must be the other red Q)

The total number of favorable outcomes is $51 * 2 * 50! = 2 * 51!$

Example: If favorable outcomes are hands of 7 cards with exactly 3 red cards, list one favorable outcome and count the number of favorable outcomes.

One favorable outcome is:

$$\{3\heartsuit, 7\clubsuit, 2\diamondsuit, 7\spadesuit, Q\clubsuit, 8\diamondsuit, J\spadesuit\}$$

In generating an example like this, I get to choose:

- which three red cards to include $\longleftarrow C(26, 3)$ options
- which four black cards to include $\longleftarrow C(26, 4)$ options

I don't get to choose:

- what order the cards go in (because a hand is a set of cards)

The total number of favorable outcomes is $C(26, 3) * C(26, 4)$

Example: If favorable outcomes are six-letter strings with two letters alternating, list one favorable outcome and count the number of favorable outcomes.

One favorable outcome is:

In generating an example like this, I get to choose:

- the first letter $\longleftarrow$ 26 options
- the second letter $\longleftarrow$ 25 options

I don't get to choose:

- the third, fourth, fifth, or sixth letters (because the alternating pattern determines what they must be)

Another equivalent way to do this is to choose:

- which set of two letters the string will include $\longleftarrow C(26, 2)$ options

- which of the two letters in the set will be first $\longleftarrow$ 2 options

This also completely determines the string.

Either way, the total number of favorable outcomes is the same, $26 * 25 = C(26, 2) * 2$.

4. Is it helpful to use the complement rule? Are there many ways in which the favorable outcome can be achieved, and only a few ways in which it's not achieved? If so, try using

$$P(A) = 1 - P(\overline{A}).$$

The words "at least" often indicate to use the complement rule.

Example: If a license place is made up of three letters followed by three numbers, like ABC-123, what is the probability of a randomly selected license plate having a palindrome in the letters or numbers?

Three ways for this to happen:

- palindrome in the letters, no palindrome in the numbers
- palindrome in the numbers, no palindrome in the letters
- palindrome in the letters and numbers

and only way for this not to happen:

- no palindrome in letters and no palindrome in numbers

To have no palindrome in letters, we need the third letter to be different from the first letter, and to have no palindrome in numbers, we need the third number to be different from the first number. So by the complement rule,

$$P(\text{palindrome in letters or numbers}) = 1 - \frac{25}{26} * \frac{9}{10} = \frac{7}{52}.$$

Example: What is the probability of at least one heads in k tosses of a fair coin?

There are many ways to have at least one heads in k tosses, but only one way not to, which is getting all tails. So,

$$P(\text{at least one heads in } k \text{ tosses}) = 1 - \left(\frac{1}{2}\right)^k.$$

Example: Every time you call your grandma, she has a $\frac{1}{3}$ chance of answering the phone. If you call your grandma three times this week, what's the probability you'll get to talk to her this week?

There are several ways in which you can talk to her, for example maybe you'll talk on the first call and third call, but not the second. There is only one way in which you don't talk to her, which is that she fails to answer the phone on all three calls. So by the complement rule,

$$P(\text{talk to Grandma}) = 1 - \left(\frac{2}{3}\right)^3 = \frac{19}{27}.$$

Even though this problem didn't use the words "at least", that is really what the question is asking - what is the probability you talk to your grandma at least once during the week?

Note also that you can't just add the probabilities for each phone call to get $\frac{1}{3} + \frac{1}{3} + \frac{1}{3} = 1$, as there is no guarantee that you'll get to talk to her. And clearly, this method has a problem if we were to call four times, since it would lead to a probability greater than one. The problem with this approach is that the cases (talk on first call, talk on second call, and talk on third call) are not disjoint. It is possible that you talk on multiple calls, and so you'd need to subtract this overlap using the addition rule. See Homework 5, problem 4 (b) and (c).

5. Is it helpful to use the multiplication rule? Do you need several things to happen all at once?

The multiplication rule says you can make the first thing happen, then you only have to worry about the second thing happening in the case that the first already happened.

$$P(A \cap B) = P(A) * P(B|A)$$

We can also use it in the other order:

$$P(A \cap B) = P(B) * P(A|B)$$

We can also use it for three or more events, by just making sure that the next thing happens, assuming that all prior things already happened. For example, for three events, the multiplication rule becomes

$$P(A \cap B \cap C) = P(A) * P(B|A) * P(C|(A \cap B)).$$

Example: Suppose you draw a random sample of 5 objects with replacement from a set of 20 distinct objects. What is the probability of having no repeated objects in the sample?

To have no repeated objects, we need several things to happen all at the same time:

- first element is distinct from all previous elements (this happens all the time)
- second element is distinct from all previous elements
- third element is distinct from all previous elements
- fourth element is distinct from all previous elements
- fifth element is distinct from all previous elements

When calculating the probabilities of each of these events, we get to assume that all the earlier events have already happened. For example, when determining the probability that the fifth element is distinct from all previous elements, we get to assume that the first four elements are already distinct from one another. This means that 4 of the 20 objects are unfavorable picks for our last element, and 16 are favorable.

Then by the multiplication rule, the probability of no repeats is

$$P(\text{no repeats}) = 1 * \frac{19}{20} * \frac{18}{20} * \frac{17}{20} * \frac{16}{20}.$$

6. Do you need to calculate a conditional probability, the probability of an event happening when you know another has occurred? The words "given that", "assuming", and "if you know" are often hints that you're being asked to find a conditional probability.

The formula for conditional probability is just the same as the multiplication rule, written in a slightly different way:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

When computing a conditional probability, you can either:

1. use the above formula, or
2. calculate the conditional probability directly.

For most problems, calculating the conditional probability directly is usually easier because using the formula requires you to find two probabilities, one for the numerator and one for the denominator, while a direct calculation is just one probability. Also, finding the probability of the intersection of two events can be harder than just finding the conditional probability.

When calculating a conditional probability directly, use the given information (the condition) to restrict the possible set of outcomes.

Example: A license place is made up of three letters followed by three numbers, like ABC-123. If you know that a randomly selected license plate starts with an A, what is the probability of that license plate having three distinct letters that appear in alphabetical order?

Try this problem both ways, using the conditional probability formula and calculating the conditional probability directly. Which way do you think is easier?

Let's try it both ways, using the formula, and calculating the conditional probability directly.

1. Using the formula,

$$P(\text{3 distinct letters in alph. order}|\text{start A}) = \frac{P(\text{3 distinct letters in alph. order} \cap \text{start A})}{P(\text{start A})}$$

   Now I have two separate probability questions to solve, one for the numerator and one for the denominator.

   For the denominator, $P(\text{start A})$, if the sample space is first letters for the license plate, then only one out of 26 possible first letters is a favorable outcome, so

$$P(\text{start A}) = \frac{1}{26}.$$

   For the numerator, $P(\text{3 distinct letters in alph. order } \cap \text{ start A})$, the sample space can be all license plates, or sequences of three letters followed by three numbers. All license plates are equally likely, and there are $26^3 * 10^3$ many.

To count the number of favorable outcomes, I'll first generate an example of a favorable outcome: ADG-557.

In generating an example like this, I get to choose:

- which two other letters besides A will be included $\longleftarrow C(25, 2)$ options
- the first number $\longleftarrow$ 10 options
- the second number $\longleftarrow$ 10 options
- the third number $\longleftarrow$ 10 options

I don't get to choose:

- the first letter (must be A)
- the order of the other two letters (must be alphabetical)

So the number of favorable outcomes is $C(25, 2) * 10^3$ and

$$P(3\text{ distinct letters in alph. order} \cap \text{start A}) = \frac{C(25,2) * 10^3}{26^3 * 10^3}.$$

Therefore, using the conditional probability formula,

$$\begin{aligned}
P(3\text{ distinct letters in alph. order}|\text{start A}) &= \frac{P(3\text{ distinct letters in alph. order} \cap \text{start A})}{P(\text{start A})} \\
&= \frac{\frac{C(25,2) * 10^3}{26^3 * 10^3}}{\frac{1}{26}} \\
&= \frac{75}{169}
\end{aligned}$$

2. To calculate the conditional probability directly, the sample space of possible outcomes will be all license plates that start with an A. Each such outcome is equally likely, and the number of such outcomes is $1 * 26 * 26 * 10 * 10 * 10$.

   To count the number of favorable outcomes, I'll first generate an example of a favorable outcome: ADG-557.

   In generating an example like this, I get to choose:

   - which two other letters besides A will be included $\longleftarrow C(25, 2)$ options
   - the first number $\longleftarrow$ 10 options
   - the second number $\longleftarrow$ 10 options
   - the third number $\longleftarrow$ 10 options

   I don't get to choose:

   - the first letter (must be A)
   - the order of the other two letters (must be alphabetical)

So the number of favorable outcomes is $C(25,2) * 10^3$, and

$$P(3 \text{ distinct letters in alph. order}|\text{start A}) = \frac{C(25,2) * 10^3}{1 * 26 * 26 * 10 * 10 * 10} = \frac{75}{169}.$$

Both ways of doing this problem require the same essential insight, which is counting the number of plates that start with A and have three distinct letters in alphabetical order, but finding the conditional probability directly is more straightforward.

7. Is it helpful to use the addition rule? Can you break down the favorable outcomes into a reasonable number of cases? Do the cases overlap? Can any one outcome fall into multiple cases? If so, don't forget to subtract off the intersection to avoid counting these outcomes multiple times.

The basic addition rule for two cases is

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

If the cases don't overlap, the last term is zero and can be left out.

For more than two cases, if the cases don't overlap, you can simply add up the individual properties of each case. If they do overlap, the formula gets complicated very quickly. You only need to know up to three cases:

$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C).$$

Example: What is the probability that a random five-letter string ends in a or z?

Since no string can end in more than one letter, the cases are disjoint.

$$P(\text{end } a \cup \text{end } z) = P(\text{end } a) + P(\text{end } z) = \frac{1}{26} + \frac{1}{26} = \frac{2}{26}.$$

Example: What is the probability that a random five-letter string starts in a or ends in z?

Now the cases are not disjoint, since there are five-letters strings like *achez* that start in a and end in z, so we need to subtract off the overlap.

$$P(\text{start } a \cup \text{end } z) = P(\text{start } a) + P(\text{end } z) - P(\text{start } a \cap \text{end } z) = \frac{1}{26} + \frac{1}{26} - \frac{1}{26} * \frac{1}{26} = \frac{51}{676}.$$

Example: Every time you call your grandma, she has a $\frac{1}{3}$ chance of answering the phone. If you call your grandma three times this week, what's the probability you'll get to talk to her this week?

We've already seen how to solve this problem with the complement rule, which is definitely the easy way to do this, however, you can also do it using the addition rule where

- A is the event that you talk to Grandma on the first call
- B is the event that you talk to Grandma on the second call
- C is the event that you talk to Grandma on the third call

The addition rule gives the same result as before:

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(C \cap B) + P(A \cap B \cap C) \\ &= \frac{1}{3} + \frac{1}{3} + \frac{1}{3} - \left(\frac{1}{3} * \frac{1}{3}\right) - \left(\frac{1}{3} * \frac{1}{3}\right) - \left(\frac{1}{3} * \frac{1}{3}\right) + \left(\frac{1}{3} * \frac{1}{3} * \frac{1}{3}\right) \\ &= \frac{19}{27} \end{aligned}$$

8. If you're counting sequences, can the sequences be organized according to templates? Templates are like cases, showing a pattern that the sequence will follow. Templates should be disjoint, not overlapping.

If each template is associated with the same probability, then each term in the addition rule has the same value so we can calculate the probability as:

$$\text{(number of templates)} * \text{(probability of each template)}.$$

Example: Suppose that every time you visit a volcano, you have a one percent chance of witnessing an eruption. If you visit five times, what is the probability of seeing your second eruption on your fifth visit?

For a second eruption on the fifth visit, you need to have seen one eruption in the first four visits. If we use Y for yes and N for no, then the templates are:

- YNNNY
- NYNNY
- NNYNY
- NNNYY

The number of templates is $C(4, 1) = 4$, to choose where in the first four visits the Y should go. The probability of the first template is

$$P(YNNNY) = 0.01 * 0.99 * 0.99 * 0.99 * 0.01 = (0.99)^3 * (0.01)^2.$$

In fact, the probability for each template is the same, since all have 2 Y's and 3 N's, so the probability of seeing your second eruption on the fifth visit is

$$4 * (0.99)^3 * (0.01)^2.$$

Example: All UCSD campus phone numbers take the form 858-534-XXXX. What is the probability of a randomly chosen UCSD phone number having each digit that's included in the ten-digit number appearing exactly twice?

Since a single 3 and a single 4 appear already, one of the last four digits must be 3 and another must be 4. The other must match and be one of $0, 1, 2, 6, 7, 9$. Letting M represent the two matching digits, the templates are:

- 34MM
- 3M4M
- 3MM4
- ...

We don't need to list them all, but we do need to be able to count how many there will be. There are 4 positions where we can put the 4 and 3 positions where we can put the 3. This completely determines the template, so the number of templates is $4 * 3 = 12$.

The probability of the first template, 34MM, is

$$\frac{1}{10} * \frac{1}{10} * \frac{6}{10} * \frac{1}{10}$$

because the first M can be any of the six digits $0, 1, 2, 6, 7, 9$, and the second M must match that. The probability of any template is the same, so the total probability of having a UCSD phone number with each digit appearing twice is

$$12 * \frac{1}{10} * \frac{1}{10} * \frac{6}{10} * \frac{1}{10}.$$

9. Can you think of an additional assumption that would make the probability easier to calculate? Can you calculate how likely that assumption is to be true? If so, you may be able to break up your probability into cases based on this extra assumption and whether it is true or not.

This approach basically uses the Law of Total Probability. If A is the event you want the probability of, and B is the extra assumption, then since B and $\overline{B}$ partition the sample space, the Law of Total Probability says

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \overline{B}) \\ &= P(A|B) * P(B) + P(A|\overline{B}) * P(\overline{B}) \end{aligned}$$

Example: Suppose that a standard deck of 52 cards is shuffled in a random order. What is the probability that both red queens are adjacent?

If we could assume that the Q$\diamondsuit$ were the first or last card in the shuffled deck, the problem would be very easy, because we'd just need to make sure that the Q$\heartsuit$ were next to it. With the Q$\diamondsuit$ on the end, the chance of the Q$\heartsuit$ being adjacent is $\frac{1}{51}$.

But this assumption is certainly not always true. It only happens with probability $\frac{2}{52}$. The rest of the time, the Q$\diamondsuit$ is in a middle position. But that case also makes the original problem much easier to solve. If the Q$\diamondsuit$ is in a middle position, the chance of the Q$\heartsuit$ being adjacent is $\frac{2}{51}$ because it could be on either side of the Q$\diamondsuit$.

So the Law of Total Probability says

$$\begin{aligned} P(\text{both red Q adjacent}) &= P(\text{both red Q adjacent}|Q\diamondsuit \text{ on end}) * P(Q\diamondsuit \text{ on end}) \\ &\quad + P(\text{both red Q adjacent}|Q\diamondsuit \text{ in middle}) * P(Q\diamondsuit \text{ in middle}) \\ &= \frac{1}{51} * \frac{2}{52} + \frac{2}{51} * \frac{50}{52} \\ &= \frac{1}{26}. \end{aligned}$$

Example: Every time you call your grandma, she has a $\frac{1}{3}$ chance of answering the phone. If you call your grandma three times this week, what's the probability you'll get to talk to her this week?

If we could assume that the first two calls went unanswered, then the probability of talking to Grandma is just $\frac{1}{3}$, which is the chance she answers your third call. The probability of the first two calls going unanswered is $\frac{2}{3} * \frac{2}{3} = \frac{4}{9}$.

If this assumption isn't true, which happens the remaining $\frac{5}{9}$ of the time, that means Grandma answered at least one of your first two calls, in which case the probability that you talk to her is 1. So,

$$\begin{aligned} P(\text{talk}) &= P(\text{talk}|\text{first two calls unanswered}) * P(\text{first two calls unanswered}) \\ &\quad + P(\text{talk}|\text{at least one of first two calls answered}) * P(\text{at least one of first two calls answered}) \\ &= \frac{1}{3} * \frac{4}{9} + 1 * \frac{5}{9} \\ &= \frac{19}{27}. \end{aligned}$$