Wednesday, September 13 **Fall 2023**

Section 1: Counting and Definitions of Probability

Srihari Ganesh

Based section template and practice problems by Rachel Li and Ginnie Ma '23

1. I am interested in five subjects (statistics, math, computer science, chemistry, and biology) and am deciding between 100, 102, 110, and 111 in each department (e.g., my options are Stat 100, Stat 102, Stat 110, Stat 111, Math 100, Math 102, etc.). Assume unless stated otherwise that each class (or set of classes) is equally likely to be chosen if possible.

1. **How many different schedules of 4 courses can I take?**

   Solution
   I select my 4 courses without replacement, and the order in which I choose the classes does not matter. This means we use the binomial coefficient with $n = 5 * 4 = 20$ and $k = 4$ to get $\binom{20}{4}$ different schedules.

2. **I want to send my friend a text of my 4-class schedule. My text will look be a list like "Chemistry 101, Stat 110, Biology 111, Math 101", which is considered a different text from "Stat 110, Chemistry 101, Biology 111, Math 101". How many different such texts could I send?**

   Solution
   This is a permutation problem: we want to pick $k = 4$ classes in order from $n = 20$ total classes. So there are $20 \times 19 \times 18 \times 17 = \frac{20!}{16!}$ possible texts.

3. **If I take 5 courses, what is the probability that I take exactly 1 course from each department?**

   Solution
   We can use naive probability since every outcome is equally likely, recalling that

$$P_{\text{naive}} = \frac{\text{\# favorable outcomes}}{\text{\# total outcomes}} = \frac{\text{\# number of schedules with one class from each department}}{\text{\# number of possible schedules with five classes}}$$

   We define S to be the sample space (possible schedules with 5 classes). We define A to be the event that the 5-class schedule has one course from each department, so

$$P_{\text{naive}} = \frac{|A|}{|S|}.$$

   Using a similar result to 1.1, number of outcomes possible outcomes is

$$|S| = \binom{20}{5} = 15504.$$

For our favorable outcomes, we can pick 1 class from each of the 4 options in each department, so

$$|A| = \left[\binom{4}{1}\right]^5 = 1024.$$

Thus, the probability of a schedule with exactly one class from each department is

$$P_{\text{naive}} = \frac{|A|}{|S|} = \frac{\left[\binom{4}{1}\right]^5}{\binom{20}{5}} \approx 0.066.$$

4. **My advisor wants to know that I have backup plans, so they ask me to send 2 schedules with 4 courses each, where the schedules can have no overlap. How many such sets of two schedules can I send my advisor?**

Solution
[Solution 1] We have $\binom{20}{4}$ ways to select the first schedule and $\binom{20-4}{4} = \binom{16}{4}$ ways to select the second schedule. However, the two schedules have no special ordering, so we have to adjust for our overcounting by dividing by 2!. This yields

$$\frac{\binom{20}{4}\binom{16}{4}}{2!}.$$

[Solution 2] We could first select the 8 classes that are in either schedule, then choose 4 classes from the 8 for the first schedule. Again, we are overcounting by labeling the "first" and "second" schedules, so we correct by dividing by 2!. This yields

$$\frac{\binom{20}{8}\binom{8}{4}}{2!}.$$

5. **After I finalize my schedule, my advisor now wants to know how many classes I'm taking in each subject, so I send them a list that follows the following format of the following example (with the subjects always in this order):**

   [Statistics: 2, Math: 0, Computer Science: 1, Chemistry: 1, Biology: 1].

   How many such lists can I send for a 4-class schedule?

   > **Solution**
   > Luckily, since there are four classes in each department, we can never run out of classes to take in any department. Thus, we can think of this problem as sampling 4 classes with replacement from 5 subjects. We can thus use Bose-Einstein with a sample of size $k = 4$ and $n = 5$ options to sample from, which means the total number of lists if
   >
   > $$\binom{n+k-1}{k} = \binom{5+4-1}{4} = \binom{8}{4} = 70.$$

6. **If I take 6 courses, what is the probability that there's at least 1 subject in which I am not taking any courses? Solve this problem in two ways.**

   > **Solution**
   > We can use naive probability since each schedule is equally likely to be selected. The total number of possible schedules is $\binom{20}{6} = 38760$, similarly to problem 1.1. Define A to be the event that I take a schedule that is missing at least one subject.
   >
   > These are the two solutions I thought of, let me know if you come up with more!
   >
   > *[Solution 1: Complementary counting]* There seem like a lot of ways to be missing at least one subject, so we can look at the complement. A^c is the event that I take a schedule including at least 1 class from every subject. Since there are 5 subjects and 6 classes, we must take exactly 2 classes in exactly 1 subject and exactly 1 class in each of the remaining 4 subjects. Thus, the number of such outcomes is given by selecting the course with two classes, $\binom{5}{1}$, selecting those two classes, $\binom{4}{2}$, and selecting one class from each of the other 4 departments, $[\binom{4}{1}]^4$, which multiplies out to
   >
   > $$\binom{5}{1}\binom{4}{2}\left[\binom{4}{1}\right]^4 = 7680.$$
   >
   > This makes $P(A^c) = \frac{7680}{38760} \approx 0.198$. Thus, $P(A) = 1 - P(A^c) \approx 0.802$.
   >
   > *[Solution 2: Principle of Inclusion-Exclusion (PIE)]* Our desired event involves missing **at least one** subject. We can thus express A as a union of A_1, A_2, A_3, A_4, A_5, where A_1 is the event that Statistics is missing from my schedule, A_2 is the event that math is missing from my schedule, etc. So $A = \bigcup_{i=1}^{5} A_i$. We can use the principle of inclusion-exclusion to

see that

$$P(A) = \sum_{i=1}^{5} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + \sum_{i<j<k} P(A_i \cap A_j \cap A_k) - \sum_{i<j<k<\ell} P(A_i \cap A_j \cap A_k \cap A_\ell) + P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$$

In event A_1, we have free pick of all subjects that are not statistics, so $P(A_1) = \frac{\binom{16}{6}}{\binom{20}{6}}$. Similarly, with $A_1 \cap A_2$ we have free pick of all subjects that are not statistics or math, so $P(A_1 \cap A_2) = \frac{\binom{12}{6}}{\binom{20}{6}}$. See that it's impossible to make a schedule with only one subject (missing 4 subjects), since there are only 4 classes per subject. Using the symmetry of the problem (i.e., the fact that each subject has the same number of classes), we know that $P(A_1) = P(A_2) = P(A_3) = \cdots$, $P(A_1 \cap A_2) = P(A_1 \cap A_4) = P(A_3 \cap A_5) = \cdots$. The calculation becomes

$$\begin{aligned} P(A) &= \binom{5}{1} P(A_1) - \binom{5}{2} P(A_1 \cap A_2) + \binom{5}{3} P(A_1 \cap A_2 \cap A_3) \\ &\quad - \binom{5}{4} P(A_1 \cap A_2 \cap A_3 \cap A_4) + P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) \\ &= \binom{5}{1} \frac{\binom{16}{6}}{\binom{20}{6}} - \binom{5}{2} \frac{\binom{12}{6}}{\binom{20}{6}} + \binom{5}{3} \frac{\binom{8}{6}}{\binom{20}{6}} - \binom{5}{4} (0) + 0 \\ &\approx 0.802. \end{aligned}$$

2. *(Problem selection and solution taken from Rachel Li and Ginnie Ma's notes)* Hockey Stick Identity: show using a story proof that

$$\sum_{j=k}^{n} \binom{j}{k} = \binom{n+1}{k+1}.$$

Hint: Imagine arranging a group of people by age, and then think about the oldest person in a chosen subgroup.

Solution
On the RHS, we can imagine we are in a room with $n+1$ people and trying to form a committee of $k+1$ people. We also specify that the oldest person on the selected committee will be designated as the President.
On the LHS, we can imagine lining up all $n+1$ people in the room by age order (with the oldest on the left and the youngest on the right). Now we can consider each person as a candidate for President. Starting on the left with the oldest person, we add this person to the committee as the President and now are left with the task of choosing k other people to be on the committee. Note that the people who are being chosen for the committee must be younger than the President. In this first case, we have n people to choose from since everyone is younger than the oldest person. That gives us $\binom{n}{k}$. Now consider the case where we add the second-oldest person to the committee as the president. To choose the k other people to add to the committee, we now have $n-1$ candidates because we cannot add the oldest person to the committee - this gives us $\binom{n-1}{k}$. We can keep going down the line until we reach the $k+1$th youngest person (who is also the $n-k$th oldest person). This leaves the k youngest people left in the room to be selected for the committee, giving us $\binom{k}{k} = 1$, which makes sense. Summing over all of these cases gives us the LHS of the identity.

3. Use problem 1.4 to construct a story proof for the following identity (assuming $n \geq 2k$). Recall that in problem 1.4, my advisor asked me to make 2 (non-overlapping) schedules with 4 courses each.

$$\binom{n}{2k}\binom{2k}{k} = \binom{n}{k}\binom{n-k}{k}.$$

Solution
The two solutions provided in problem 1.4 make up the story proof. Essentially, with $n = 20$ possible classes and the task of making two non-overlapping schedules with $k = 4$ classes each, we can either select each set of $k = 4$ individually, or select all $2k = 8$ classes and split them up afterwards. This gives

$$\frac{\binom{n}{2k}\binom{2k}{k}}{2!} = \frac{\binom{n}{k}\binom{n-k}{k}}{2!}.$$

Canceling the denominators gives completes the story proof.