Name ____________________

CS200 Worksheet - Counting and Probability

1. How many DNA sequences of length 6 ($\{G, A, C, T\}^6$) (you do not need to simplify your answer)

   (a) Do not contain T?

   (b) Contain the base A?

   (c) Do not contain all 4 base pairs?

   (d) Contain all 4 base pairs?

   (e) Contain the ordered sequence CAT? (Careful of overcounting.)

   (f) Contain exactly 2 of the 4 base pairs?

   Solution

   (a) We have 3 choices (G,A,C) for each of the 6 positions, so using the product rule there are $3 \times 3 \times 3 \times 3 \times 3 \times 3 = 3^6$ sequences.

   (b) The number that contain the base A is equal to the the total number of sequences, minus the number that do not contain A. The total number is 4^6 by the product rule, and the number that do not contain A is 3^6, so the total is $4^6 - 3^6$.

   (c) See next part: $4^6 - \binom{4}{1}\binom{6}{3}3! + \binom{4}{2}\binom{6}{4}\binom{4}{2}2!$

   (d) There are two options - we could have one letter appear 3 times, and the rest appear once, or we could have two letters appear two times each, and the other two appear once. Let's count the number of strings with the first option. We first find the number of ways we could choose which letter to appear 3 times. There are $\binom{4}{1}$ ways of doing this. Next we choose the positions of those three appearances: $\binom{6}{3}$. Finally, there are 3! ways of permuting the final three letters in the remaining three spots. This gives $\binom{4}{1}\binom{6}{3}3!$ ways if one letter appears 3 times. Next suppose two letters each appear twice. There are $\binom{4}{2}$ ways of picking these two letters. Now there are $\binom{6}{4}$ ways of choosing where these four letters will appear in the string. Next among those four position, we need to choose two to put the first pair, and the second pair will go in the next pair. There are $\binom{4}{2}$ ways to do this. Finally, there are 2! ways to permute the final two letters. This gives us $\binom{4}{2}\binom{6}{4}\binom{4}{2}2!$. Thus the total number of strings is $\binom{4}{1}\binom{6}{3}3! + \binom{4}{2}\binom{6}{4}\binom{4}{2}2!$

   (e) There are 4 positions where CAT could be: CATxxx, xCATxx, xxCATx, xxxCAT. Then for each of these options, we have to choose one of the four letters for each of the other three positions, so there are 4^3 ways of choosing the other three letters. So using the product rule, we have 4×4^3 sequences. But then we've overcounted the sequence CATCAT. The final number of strings is thus $4 \times 4^3 - 1$.

(f) First we count the number of ways of choosing 2 out of 4 base pairs. There are $\binom{4}{2}$ choices for which two base pairs to include. Once we've chosen which two base pairs to include, there are 2^6 sequences using only these two base pairs. But there are two of these sequences that only contain 1 base pair (the case where all 6 letters in the sequence are the same), so there are $2^6 - 2$ that contain two base pairs. Thus the total number is $\binom{4}{2}(2^6 - 2)$.

2. Explain why there are at least six people in California (population: 37 million) with the same three initials who were born on the same day of the year (but not necessarily in the same year). Assume that everyone has three initials and there are 366 days in the year, including leap years. (Remember the pigeon hole principle!)

Solution There are 366 possible birthdays, and 26^3 possible initials, so there are $366 \times 26^3 = 6432816$. Now suppose for contradiction that no 6 people have the same birthday and same initials in California. This means at most 5 people have the same birthday and same initials. But $5 \times 6415240 \approx 32$ million, which is not enough. So there must be at least 6 people that have the same initials and birthday.

3. A coin is flipped 10 times. If the order of the outcomes matters, what is the probability that

   (a) The sequence alternates heads and tails?

   (b) There are exactly 5 heads and 5 tails?

   (c) There are at most 3 tails?

Solution For all of these problems, the size of the sample space is 2^{10} (using the product rule).

(a) There are only two ways to do this: start with a heads or tails. Thus the probability is $2/2^{10} = 2^{-9}$.

(b) There are $\binom{10}{5}$ ways of getting exactly 5 tails, so the probability is

$$\frac{\binom{10}{5}}{2^{10}} \tag{1}$$

(c) There are 4 ways we can have at most 3 tails: getting 3 tails, 2 tails, 1 tail, or zero tails. There are $\binom{10}{3}$ ways of getting 3 tails, $\binom{10}{2}$ ways of getting 2 tails, $\binom{10}{1}$ ways of getting 1 tails, and one way of getting 0 tails. Thus the probability of getting at most 3 tails is

$$\frac{\binom{10}{3} + \binom{10}{2} + \binom{10}{1} + 1}{2^{10}} \tag{2}$$

4. In roulette, you spin a ball on a wheel to get one of 38 possible numbers. There are 18 black numbers, 18 red numbers, and 2 white numbers.

(a) What is the probability that you get a red number?

(b) What is the probability that you get a black number twice in a row? What is the probability that you get a white number at any point in 5 spins.

Solution

(a) The size of the sample space for one round is 38. Then there are 18 red outcomes. So the probability is 18/38.

(b) Since we have two rounds, using the product rule, the size of the sample space is 38^2. The number of ways that we can get a black number twice is 18^2 using the product rule. Thus the probability is $18^2/38^2$.

(c) We first calculate the probability that you don't get a white number in 5 spins. The size of the sample space is 38^5. Then the ways you don't get a white number in each spin is 36, so the probability that you don't get a white number is $36^5/38^5$. So the probability that you get a white number is $1 - 36^5/38^5$.

5. Suppose you have a loaded di, where a 3 is twice as likely to appear as any of the other five outcomes.

   (a) What is the probability of each outcome?

   (b) What is the probability of getting an even outcome?

Solution

(a) Let p be the probability of getting an outcome that is not 3. Then

$$1 = \sum_{i \in \{1,2,3,4,5,6\}} Pr(i) = 2p + \sum_{i \in \{1,2,4,5,6\}} p = 7p, \tag{3}$$

so $p = 1/7$, and $Pr(3) = 2/7$.

(b) The event we are interested in is $E = \{2, 4, 6\}$. Then

$$Pr(E) = \sum_{i \in E} Pr(i) = 3/7. \tag{4}$$