CHEM 116 January 25, 2019

Unit 2, Lecture 1

Numerical Methods and Statistics

Companion Reading

Bulmer Chapter 3

1 Probability II

1.1 Combinations and Permutations

We'll take a quick detour and learn about counting, or combinatorial math.

If you have n objects and are making samples of length l, there are n^l permutations

This one is quite simple. Think about a number. You know the number of digits places and 0-9 can go in each. So l is the number of digits places and n is the number of possible numbers. For example, there are 10^4 possible 4 digit numbers (0000-9999), as expected.

If you have n items and it happens that you're making samples of length n, the number of permutations is $n!$.

Consider the three letters in the word 'cat'. You can rearrange this many ways:

1. cat
2. cta
3. tac
4. tca
5. atc
6. act

See the pattern of three possibilities for the first letter. Then, for each first letter there are two possibilities. Then knowing the first two letters, there is only one possibility for the last: $3 \times 2 \times 1$.

if you have n items and are making samples of length l, but you disallow repeats of the n items you have this many permutations:

$$\frac{n!}{(n-l)!}$$

This is similar to the equation above, but our factorial starts at n and counts down until we've exhausted l. For example, if we have the letters A, B, C, D and E and are making all 3-letter combinations, we have: $5 \times 4 \times 3$. That's what the equation above is.

if you have n items and are making samples of length l, but you disallow repeats of the n items you have this many combinations:

$$\frac{n!}{l!(n-l)!} = \binom{n}{l}$$

This is the same as above, but divided by $l!$ since permutations of the same combination are not counted twice (e.g., 1, 2, 3 is the same combination as 3, 2, 1). How did we know there are $l!$ permutations for each combination? From the rule above.

if you have n items and are making samples of length l, and you allow repeats of the n items you have this many combinations:

$$\frac{(n+l-1)!}{l!(n-1)!}$$

This one is a little harder to understand. Look up "Stars and Bars" which is a well-known derivation of this formula. *One important detail is that we can now have $l > n$ or $l < n$ since we allow repeats.* Let's see an example. Consider the letters D, F, G and I'm making combinations of length 2:

1. DD
2. FF
3. GG
4. DF
5. DG
6. FG

Notice we don't have FD in our list because that's the same combination as DF. Using our equation above we get

$$\frac{(3+2-1)!}{2!(3-1)!} = \frac{4 \times 3 \times 2 \times 1}{2 \times 1 \times 2 \times 1} = \frac{24}{4} = 6$$

For comparison, the number of permutations here is $3^2 = 9$. The three items that differentiate the permutations and combinations are FD, DG, GF.

We know now all the possible cases for combinations and permutations.

1.1.1 More Examples

Let's say my sample space is 5 digit numbers. How large is it?

$$10^5$$

Given that there are four possible pets (cats, foxes, dogs, horse) and everyone owns 2, what's the sample space size of pet ownership?

If we disallow multiple pets of the same type: $\frac{4!}{(4-2)!} = 12$. Or $4^2 = 16$

What's the sample space size of a shuffled deck of cards?

$$52! \approx 10^{68}$$

for reference there are 10^{57} atoms in the solar system

What's the sample space size for a draw of 5 cards?

$$\frac{52!}{5!(52-5)!} = 2,598,960$$

How many combinations of rolls of 2 dices are possible?

$$\frac{(6+2-1)!}{2!(6-1)!} = 21$$

1.2 Repeated Items in Permutations

What happens when one of our n items has a repeat? For example, consider rearranging the letters in the word 'DAD':

1. DAD
2. DDA
3. ADD

However, our previous rule tells us that there should be $3! = 6$ permutations. The reason why is that we 'lose' some permutations because the two Ds are interchangeble. For each of the words we could swap the two Ds and keep the same word. Thus for each two permutations according to our original rule, there is actually only one unique permutations:

$$\frac{3!}{2} = 3$$

Let's now look at three repeated items: 'ABBB':

1. ABBB
2. BABB
3. BBAB
4. BBBA

We see 4 permutations and our rule says $4! = 24$. To see what's happening, consider one of our unique permutations but label the letters:

- $\mathrm{AB_1B_2B_3}$
- $\mathrm{AB_1B_2B_3}$
- $\mathrm{AB_3B_2B_1}$
- ...

Thus for each of our unique permutations, there are 3! ways to rearrange but keep the same permutation. Put another way, of the 24 rearrangements, we over-count by 3!.

Following similar logic you can see that if we have multiple repeated items, the product of their repeats shows up in the denominator. For example, if you have 'ABBBDDCE', the number of permutations is:

$$\frac{8!}{3! \times 2!} = 3360$$

Symbolically this is:

$$\frac{n!}{\prod_i k_i!} \tag{1}$$

1.3 Multiple Combinations

Consider if we don't specify l for combinations. For example, say you have n items and you can make all combinations from $l = 0$ through $l = n$. To analyze this, we can just sum our rules for each l:

$$\sum_{l=0}^{n} \frac{n!}{l!(n-l)!}$$

This expression can be simplified:

$$\sum_{l=0}^{n} \frac{n!}{l!(n-l)!} = 2^n \tag{2}$$

This can be thought of as the number of subsets you can make. Subsets, like we saw for events, are sets of sets. For example if you have a set consisting of $\{3, 5, 2\}$, the possible subsets are:

1. $l = 0$, $\{\}$
2. $l = 1$, $\{3\}$
3. $l = 1$, $\{5\}$
4. $l = 1$, $\{2\}$
5. $l = 2$, $\{2, 3\}$
6. $l = 2$, $\{2, 5\}$
7. $l = 2$, $\{3, 5\}$
8. $l = 3$, $\{2, 3, 5\}$

So there are $2^3 = 8$ combinations, including the 'empty' combination. Notice that when $n = l$, there is one combination. Don't get confused: when $n = l$ there is 1 combination and $n!$ permutations.

1.4 Multidimensional Sample Spaces

Nearly all examples in class thus far used sample spaces that have a single dimension. Sample spaces can be pairs of values in a 2D space or higher.

- the sample space could be the roll of two dice
- The sample space could be GPS coordinates

Sample spaces can be joined together into **product spaces**, indicated as $Q_1 \otimes Q_2$. A product space is every possible pairing of two spaces, meaning the number of elements is the product of the number of elements of the component spaces ($|Q| = |Q_1| \times |Q_2|$).

- Color of my shirt by weather today
- 5 tempertaure sensors
- Concentration of reactant A, B and product C
- The roll of two die is a product space, 6×6
- The roll of n-die is 6^n
- $\{-1, 0, 1\} \otimes \{0, 1\} = \{(-1, 0), (0, 0), (1, 0), (-1, 1), (0, 1), (1, 1)\}$

Note we are not observing two samples now where we previously observed one. Instead we are observing samples composed of multiple values (tuples).

1.5 Event vs Sample on Product Spaces

Is rolling a 5 a sample or event? What if our space is rolling two dice? Either $n = 1,\ Q = \{1, \ldots, 6\}$ or $n = 2 \times 6 - 1,\ Q = \{1, \ldots, 6\} \otimes \{1, \ldots, 6\}$. 1/6 vs 11/36.

What about weather by day of the week. If it's snowing is now an event. Why?

$$\mathrm{P}(\text{snowing}) = \mathrm{P}(\text{snowing and Monday}) + \mathrm{P}(\text{snowing and Tuesday}) + \ldots$$

2 Random Variables

A random variable is a *function* whose domain is a sample space and whose range is continuous or discrete values. Because a random variable is defined in a sample space, we may compute probabilities for each of its possible outputs.

2.1 Examples of random variables

- On the sample space of the roll of a die, the value of the sample (1,2,3,4,5,6) is an rv.
- On the sample space of the roll of a die, the square of the sample (1,4,9,16,25,36) is an rv.
- On the pathway example from homework, the length of a path is the rv.
- On the sample space of temperature, the value of temperature is an rv.
- On the sample space of temperature, an indicator of 0,1 if the temperature is above a set value is an rv.

2.2 Writing random variables

Random variables are written like X and the probability that an rv takes a specific value, x is written as $\mathrm{P}(X = x)$. This is called a **probability mass function**. The reason we introduce x is that the right-hand-side typically depends on x. For example, consider our biased die from last lecture where the relative probabilities of a roll follow the value of the roll (e.g., a 2 is twice as likely as a 1). If the rv is the value of the roll:

$$\mathrm{P}(X = x) = \frac{x}{21}$$

if the rv is the square:

$$\mathrm{P}(Y = y) = \frac{\sqrt{y}}{21}$$

$$\mathrm{P}(Y = 25) = \frac{\sqrt{25}}{21} = \frac{5}{21}$$

Almost always, we will omit the X in the probability mass function and only write $P(x)$ instead of $P(X = x)$

2.3 Random Variables in Multidimensional Sample Spaces

This adds complexity in the rv on the sample space. Let's take the sample space of lattitude and longitude. We could define two rvs, X and Y that are the lattitude and longitude, respectively. However, we could also define Z which is the product of lattitude and longitude.

Do not confuse the dimension of the sample space and the number of rvs. They are not the same.

2.4 Complexity of Continuous Random Variables

We are measuring the concentration of a reaction, θ and the sample space is 0 to 20 M. We note a value, 2.55. What is $P(\Theta = 2.55)$? Discuss

We will never ever use probability mass functions with continuous random variables. It makes no sense. Instead, we always deal with intervals. For example, what is the probability of the concentration being between 2.549 and 2.501? That can be computed using the **probability density function** (PDF). The PDF may be thought of as the pseudo-derivative of a probability mass function. It's used like so:

$$\mathrm{P}(2.549 < x < 2.501) = \int_{2.549}^{2.501} p(x)\, dx$$

Notice an integral is used to treat intervals and the PDF ($p(x)$) is indicated with a lower-case.

We will never deal with single points in continuous rv and you should always be thinking in terms of intervals over the PDF instead.

Example: Uniform distribution

$$p(x) = 1$$

due to normalization

$$\int_L^U p(x)\, dx = 1, \Rightarrow p(x) = \frac{1}{U - L}$$