CHEM 116 January 21, 2016

Unit 2, Lecture 1

Numerical Methods and Statistics

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.

1.1.1 Examples

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

$$10^5$$

Given that there are three 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?

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

1.2 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 $Q_1 \times Q_2$. 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.

- 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
- $\{-1, 0, 1\} \times \{0, 1\} = \{(-1, 0), (0, 0), (1, 0), (-1, 1), (0, 1), (1, 1)\}$

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

1.3 Event vs Sample on Product Spaces

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

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

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

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 HW 1, 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

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

3 Equations with Random Variables

3.1 Joint Probability Distribution

We will indicate the probability that two rvs, X and Y, adopt a particular value x and y *simultaneously* as $\mathrm{P}(x, y)$. This is called a joint probability distribution. Joints indicate simultaneous occurance, unlike our **AND** from last lecture.

To treat successive observations, we simply rearrange our current definitions. Take flipping a coin. We redefine our sample space to be the product space of trials flip 1 and flip 2, so (H, H), (H, T), (T, H) and (T, T) where H =heads, T =tails. Next, we say X is the observation of trial 1 and Y is the observation of trial 2.

The JPD is more flexible than our **AND** since it treats successive and simultaneous.

The continuous analogue is $p(x, y)$, which is not meaningful unless integrated over an area. Example: observing particle in fixed area.

3.2 Marginal Probability Distribution

The marginal probability distribution function is defined as

$$\mathrm{P}(x) = \sum_y P(x, y) \tag{1}$$

The marginal means the probability of $X = x, Y = y_0$ or $X = x, Y = y_1$ or $X = x, Y = y_2$, etc. For example, if X is the weather and Y is the day of the week, it is the probability of the weather 'averaged' over all possible weekdays.

The marginal allows us to remove a rv or sample dimension. That process is called **marginalizing** and the resulting $\mathrm{P}(x)$ is called the marginal. Marginalization works even if the two pieces of the joint are not independent.

3.3 Conditional Probability Distribution

A conditional probability distribution allows us to fix one sample, event, or rv in order to calculate another. It is written as $\mathrm{P}(X = x|Y = y)$. For example, what is the probability of having the flu given I'm showing cold/flu symptomps. Conditionals are generally much easier specify, understand and measure than joints or marginals.

- The probability a reaction being complete given that the temperature is less than 750 K

- The probability I visited node C given that I started in node A and ended in node B
- The probability a test shows I have influenza given that I do not have influenza (false negative)
- The probability I rolled a 7 given that one die shows 4.

The definition of a conditional probability distribution is

$$\mathrm{P}(x|y) = \frac{\mathrm{P}(x,y)}{\mathrm{P}(y)} \tag{2}$$

A CPD is a full-fledged PMF, so $\sum_{\mathcal{X}} \mathrm{P}(x|y) = 1$ due to normalization, sometimes called the law of total probability. If we forget what goes in the denominator on the right-hand-side we can quickly see that $\sum_{\mathcal{X}} \mathrm{P}(x,y)/\mathrm{P}(y) = 1$ whereas $\sum_{\mathcal{X}} \mathrm{P}(x,y)/\mathrm{P}(x) \neq 1$.

The definition is the same for continuous distributions.

This leads to an alternative way to marginalize:

$$\sum_{\mathcal{Y}} \mathrm{P}(x|y)\mathrm{P}(y) = \sum_{\mathcal{Y}} \mathrm{P}(x,y) = \mathrm{P}(x)$$

3.4 Viewing Conditionals as Sample Space Reduction

Consider guessing binary numbers at random between 0 and 7:

000
001
010
011
100
101
110
111

The probability of sampling 4 (100),

$$\mathrm{P}(x = 100) = \frac{1}{8}$$

Now, consider the rv Y, the number of non-zero bits. What is

$$\mathrm{P}(x = 100|Y = 1)$$

We could rewrite this in terms of the joint and marginal, as

$$\mathrm{P}(x = 100|Y = 1) = \frac{(x = 100, Y = 1)}{Y = 1} = \frac{1/8}{3/8} = \frac{1}{3}$$

Or we could recognize that the condition of $Y = 1$ reduces the sample space to 3, because there are only 3 samples that are consistent with $Y = 1$. Thus, the probability of $x = 100$ is $1/3$, since $x = 100$ has a single permutation and Q_c, the conditional sample space is 3.

4 Tricky Concepts

Product Spaces A product space is for joining two possibly dependent sample spaces. It can also be used to join sequential trials.

Event vs Sample on Product Spaces Things which were samples on the components of a product space are now events due to permutations

Random Variables They assign a numerical value at each sample in a sample space, but we typically care about the probability of those numerical values (PMF). So X goes from sample to number and $\mathrm{P}(x)$ goes from number to probability.

Continuous PDF A pdf is a tool for computing things, not something meaningful by itself.

Marginal Probability Distribution A marginal 'integrates/sums' out other samples/random variables/events we are not interested in.

Joint vs Conditional People almost never think in terms of joints. Conditionals are usually easier to think about, specify, and be a way to attack problems.