CHEM 116 January 17, 2016

Unit 1, Lecture 1

Numerical Methods and Statistics

1 Course Structure

1.1 Goal of course

Real world problems aren't perfect, they require statistics. Real world problems require computers. This course will equip you for that.

1.2 Components of Course

Probability Theory: The analysis of randomness of sets and variables.

Statistics: The analysis of data using probability theory

Numerical Methods: The solution of numerical problems with algorithms

1.3 Course Textbooks

Principles of Statistics A book concerned with the theory of statistics. It has hilarious examples and seems to be set in the Charles Dickens universe

Practical Statistics Explained A practical handbook concerned with the application of statistics

Code Academy A simple introduction to python: www.codecademy.com/learn/python

Python in Easy Steps Accessible book covering Python basics. Bright colors, friendly cartoons that talk to you in the margins.

1.4 Grading

This course is primarily electronic so homework and projects are more heavily weighted.

1.5 Policies

See syllabus

1.6 Example problems

- Flow on graphs (electricity, chemical reactions, fluids)
- Fitting data to kinetic rate laws
- Solving equations numerically ($\sin(x) = x$)
- Create a website containing equations, graphs and text for distributing work

1.7 Projects

Will be discussed more in the future.

1.8 Python

1. Good introductory language, used by 8/10 top CS departments
2. Wide use out of industry
3. Less used than MatLab and Excel, but growing
4. Many libraries in engineering and significantly more than MatLab outside of engineering
5. These libraries allow us to mix cool things together, like live websites with instrument data
6. Programs will be easy distribute, easy to install, easy to view by people with no programming experience

2 Probability Theory

2.1 Sample spaces

Sets, ordered sets, integers, real numbers.

Examples:

1. $\{A, B\}$ (set)
2. The roll of a die: $1 - 6$ (integers, which is an ordered set, which is a set)
3. Cards in a 52 card deck (ordered set, which is a set)
4. All possible molecules formed in a chemical reaction (set)
5. Electronic configurations of a molecule (set, possibly ordered with energy)
6. Flow rate into a tank (real number)
7. Temperature (real number)
8. Value of cryptographic key (integer)
9. Pet type owned (cat, dog, fox) (set)

Think about what is ordered, what is a set, what are real numbers, etc

2.2 Probability of sample spaces

Probability assigns a number P to each sample x. The only requirement is that the sum of $P(x)$ over the sample space is 1. That condition is called normalization.

Notice that even if $\sum_X P(x) > 1$, as long as it's finite we can *normalize* the probability by dividing it by a constant such that its sum is 1. Example:

$$P(\text{die roll}) = \text{value of die} \tag{1}$$

is not normalized, because $\sum_X P(x) = 21$. Normalizing it gives:

$$P(\text{die roll}) = \frac{\text{value of die}}{21} \tag{2}$$

2.3 Probability Algebra - For Samples

OR

The probability of sample A or sample B being drawn is exactly:

$$P(A \,\text{or}\, B) = P(A) + P(B) \tag{3}$$

AND

If we draw two samples sequentially (!) and independently (!):

$$P(A \,\text{and}\, B) = P(A)P(B) \tag{4}$$

For now, independence between trials means the outcome of one doesn't affect the probability of the other.

NOT

$$P(\sim A) = 1 - P(A) \tag{5}$$

These statements allow us to bridge probabilities together:

```
Draw an ace of spades AND roll a 2 OR roll a 4 AND NOT own a cat
```

Notice that AND is used to bridge together independent samples, OR is used to bridge together multiple possibilities, and NOT is used to "invert" probabilities.

What is wrong with this statement?

```
Draw an ace of spades OR roll a 2
```

You cannot join samples which are from different sample spaces

2.4 Events

Events are collections of samples occurring. Events can overlap, occupy entire sample space or be generally messy.

- Roll an odd number
- Have a flow greater than 2 kg/s
- Have a flow between 2 and 4 kg/s

AND and **NOT** apply to probabilities of events. **OR** only applies if the events do not overlap (mutually exclusive). For example if event A is roll an odd number and event B is roll a 3, event A includes event B meaning the normal **OR** does not apply. To deal with this, generally you redefine your event or compute the intersection $P(A \cap B)$ and use:

$$P(A \,\text{or}\, B) = P(A) + P(B) - P(A \cap B) \tag{6}$$

I hesitate to say $P(A \cap B)$ means probability of A and B occurring, because that sounds similar to what was above. Think instead that $P(A \cap B)$ is the probability of A and B occurring simultaneously. Alternatively, $P(A \cap B)$ is the sum of probabilities of the elements in the sample space that overlap in A and B.

2.5 Independent samples and combination vs permutation

When sampling multiple independent trials/events/samples, the **AND** rule applies to permutations. A **combination** of events have no ordering - order does not matter. A particular sequence from those events is called a **permutation** - order matters

Sometimes we do not care about order, like when rolling two dice: $3, 2$ is the same as $2, 3$. There are two permutations, and we must consider both to get the probability of the whole 2 observation combination. This can be done with the **OR** rule:

$$P(3, 2) = P(3) \times P(2) + P(2) \times P(3) = 2 \times P(3) \times P(2) \tag{7}$$

So, to get the probability of a permutation we can just use the **AND** rule. To get the probability of a combination, we must use the **AND** rule for each permutation and use the **OR** rule to combine each permutation that is possible for the combination. For example, a combination of 1,2,3 is the sum of the probabilities of each possible permutation of rolling a 1,2,3:

1. $P(\text{perm} : 1, 2, 3) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$
2. $P(\text{perm} : 1, 3, 2) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$
3. $P(\text{perm} : 2, 3, 1) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$
4. $P(\text{perm} : 2, 1, 3) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$
5. $P(\text{perm} : 3, 1, 2) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$
6. $P(\text{perm} : 3, 2, 1) = P(1) \cdot P(2) \cdot P(3) = \frac{1}{6^3}$

So the probability of the combination of 1,2,3 is $\frac{6}{6^3}$.

2.6 Uniform Probability Sample Spaces

If all samples have equal probabilities probabilities, the probability of a sample is

$$P(x) = \frac{1}{Q} \tag{8}$$

where Q is the sample space size.

The probability of an *event* occurring is the number of samples in the event, q, divided by the size of the sample space, Q:

$$P(\text{event}) = \frac{q}{Q} \tag{9}$$

For example, the probability of rolling an odd number is $3/6$.

The probability of a *combination* occurring is then the number of permutations, n, times the probability of a single permutation:

$$P(\text{combination}) = nP(\text{permutation}) \tag{10}$$

See the above permutation example to see this. Notice this is about multiple observations, whereas the previous two equations are about single observations.

2.7 Tricky concepts to review

Independence : For now it means the trials don't affect one another. One way to tell is if the trials can be permuted, they are independent.

OR : Cannot combine statements in different samples spaces, whereas **AND** can cross sample spaces.

Normalization : As long as your made-up probability measure is finite everywhere, it can be normalized.

Combination vs Permutation : A permutation is a particular ordering of a combination.

Event vs. Observation : I've used observation here to indicate the generation of a sample (outcome). The generation of the sample may correspond to an event, but don't confuse the word 'event' with meaning we generated a sample. An event is a set of samples, where if any elements of the set occur then we say the event occured. An example of an event is the set of $1, 5$ for rolling a die. The observation, or generation of a sample, could be 1 for example.