Week Nine Reading Notes

Notes for Lecture Sequence 16

By now, you should appreciate the ability to programmatically control a plot, how it looks, what and where the labels are placed. Graphical user interfaces are almost always more limited than programmatic ones. Being able to program gives you special powers! Remember to simply smile when someone asks you what version of Excel you used to make some particular unusual, but effective, Pylab plot.

Humans like to model or abstract their world, and lots can be learned from a detailed computer simulation of some aspect of our world. Fortunately, randomness makes these simulations easier by not having to program every last detail. Simulations rely on randomness, and so we need to understand it from a computer science perspective. Be thankful that we do not need to prove theorems but just make use of the concepts and techniques pioneered by mathematicians and statisticians.

For a long time statistics was a topic that occurred in many disciplines. The social scientists take polls of samples to see how populations differ or change over time. Doctors use statistics to see how well drugs perform, and economics use statistics to understand how economies are performing. Of course, everyone uses computers. The world is rapidly changing, and now there is **big data** – it is estimated that 2.5 quintillion bytes of data (1 followed by 18 zeros!) is produced by internet users every day[1]! Suddenly, statistics and machine learning has moved firmly into the area of computer science.

Be careful with the terminology. The concept of ‘standard error’, which is explained at the end of the first video segment, is not indenting too few columns, but it is rather a statistical property. It is an easy formula, but as often happens in live classroom situations, information presented in the last few moments of a lecture is often lost. Pay close attention to the full lecture, including the end, to make sure you don’t lose out on any info!

My favorite explanation of Monty Hall Problem is as follows:

Modified Monty Hall Problem: There are 100 doors, 99 goats, and 1 car. Choose a door, any door. You’ll get to see 98 open doors; that is, the location of 98 of the goats is revealed to you. Two doors remain closed – the door you picked, and another door.

Choose a door to open to reveal your prize. Do you stay, or do you switch?

Case A: If the door you originally picked has a goat, then there are 98 goats left. You will see all the other 98 goats. The remaining door has the prize. So you should switch.

Case B: If the door you originally picked has the prize, then there are 99 goats left. You will see 98 goats. The remaining door has a goat. So you should stick.

How often does case A occur? 99/100 times
How often does case B occur? 1/100 times

Thus, 99/100 times it is worth switching, and 1/100 times it is worth sticking.

---

[1] http://www-01.ibm.com/software/data/bigdata/

<u>Notes for Lecture Sequence 17</u>

Our goal is that you will get a feel for what computers and computation can do for you, and not for you to be an expert in all areas of computer science. The past several lectures have highlighted the power of randomness. It is a pleasant surprise to see how randomness helps understand our world.

Lecture 17 takes a slightly different approach by considering how experiments help us understand our world. Experimental data is collected and useful in many fields. At MIT, nearly all students have taken a basic course in physics and are comfortable with the basics of mathematics. Do not worry if you have forgotten your physics, logarithms, or other mathematics. The lectures do not assume any specific knowledge and everything is explained. On the other hand, we do make one big assumption about your abilities. We assume that you are willing to listen, follow along, and do not suffer from math anxiety. MIT students do not have math anxiety and while taking this course, you are an MIT student.

A very neat feature of computer science is how well it interacts with other fields. The lecture talks about how to use computation to find a curve that fits experimental data. It then takes the next step to use the values of the curve that best fits the data to create a model. Finally, that model can be used to predict the results of future experiments.

When the Apollo mission returned to earth with moon rocks, NASA deliberately withheld some of the rocks from researchers. At first it might seem wrong to withhold evidence that could help scientists. But really, NASA was helping scientists develop theories, because a theory is no good unless it can correctly predict something unknown.

Scientists developed theories based on the rocks they were given access to. NASA was happy to verify predictions about the rocks that it had held back from the scientists. If NASA had given the scientists all the rocks, it would be much harder to verify their predictions.

The lecture is not about moon rocks, but it is about springs and arrows. Pay careful attention to the lectures and try to guess what Professor Terman said that was mis-transcribed as "parse x". While you are at it, can you tell from the data if the arrow hit the bullseye?

-Larry Rudolph