Class Outline

- Ask how is everyone today: Talk to them about Science Popularization (Startalk, Numberphile, Cosmos);
- Lecture Dates: No class next week (29th). In the other week, the class will be on Friday (9th);
- Homework: Show of hands for homework. Ask for a few people to introduce their research, ask about experiments – how do you validate your research? Write on the blackboard ways for validating their research;
- Examination: Remind them that the final evaluation will be based on a work where they have to make an experiment design in the context of their own research;

Class Outline

- Define the Class Outline for Today: Today we will be describing how to prepare an experiment in computer sciences. Our focus will be the idea of New Experimentalism, a series of best practices for analysing scientific ideas based on experimental findings.
- Class resource: A very large part of today's class is based off the second and third chapter to “Experimental Methods for the Analysis of Optimization Algorithms”, by Thomas Bartz-Bielstein et al. Read it for more info!
- Review the last class briefly: The conclusion is that today we will talk about good practices when preparing a scientific activity.

Experimentalism

- Experimentalism: Learning from the observation of data that arises from experiments.
- It is very common to do this in a hapzard manner, but it is important to give a good structure to the experiment. A well structured experiment gives results that reliable, and we can try to specify what kind of results that we want.
- Definition of Factors (Factors are quantities that change in an experiment;
- Computer Scientists (unlike say, agriculture) can control most, if not all, factors in an experiment; Therefore, it is possible to make much more meaningful experiments – and also much more meaningless experiments as well.

Experimentation in computer science

- The good old days: In the good old days, running two algorithms and comparing the averages was enough. However, we have almost total control over our experiment environment, so we can do much better than that.
- One point is the use of better statistics (we will talk about that later).
- But before we even get to discuss better statistics, we need to think about how we approach the decision of making an experiment.
- This is the idea of experiment design.

Characteristics of Computer Science experiments

- Before anything else, we need to ask Why are we doing this experiment?
- These are frequent reasons for performing experiments in Computer Science:
  - Are we trying to make our algorithm solve a problem that was too hard before? Are we trying to solve a harder version of an old problem, or a new problem that has not been solved before?
  - Are we trying to improve the performance compared to the current best algorithm? Maybe making it faster, or use less memory?
  - Are we trying to understand WHY the algorithm works? What makes it work faster or slower? What are the factors that have the biggest influence in the correctness or performance of an algorithm?
- When we define why we want to do an experiment, we can decide what the best experiment is. (some people use the same experiment for everything)

Pre-experimental Setup

(Improving the performance experiments)

- Understanding the goals of the experiment will allow you to determine the kind of experimentation to make, and what factors to isolate during experimentation.
- Should we focus on Algorithm Parameters? (Algorithm tuning) - Population size, number of layers in a perceptron model, number of nodes in a parallel environment. Can you give other ideas?
- Should we focus on Problem Parameters? (Problem design) - Data set to be used, number of dimensions in an artificial data set. Number of constraints to be taken into account, or to be ignored.
- Use of real world problems vs Theoretical problems. Real world settings usually allow for little variation in the Problem-related factors, while theoretical problem settings allow for a huge control of factors.

Pre-experimental Design

(Understanding the algorithm-kind of experiments)

- We are more interested in the effect of changing the factors in the outcome, than the outcome itself.
- Because of this, we need to stablish a “policy” for varying the parameters in a controlled fashion.
- Uncertainty Analysis - for a fixed set of factors, what is the expected variation in the performance/behavior of an experiment?
- Sensitivity Analysis - now that we know the variance, which are the factors that contribute most to this variance?
- Structure of The experiment - Observation of the System vs Planned Interference;

Experimentation Models

- In the context of experimentalism, the idea of establishing Scientific Models is crucial for defining the best experiments to be conducted;
- Three levels of scientific models: Models of Scientific Hypothesis, Models of Experiment, Models of Data;
- Scientific Models are used to break a scientific inquiry into one or more “decision points” that can be analysed in isolation.
- Experimental Models transform the scientific models into practical questions that can be answered by using experiments;
- Data Models are used to generate and model the data obtained FROM the experiments, from which the above models are judged.
- Show the example with the image and the tree of models;

Experimentation Issues

- Floor and Ceiling effects: problem is too easy, problem is too hard, N is too big, N is too small;
- Confounded Effects: Factors are not perfectly isolated, We cannot change all factors at the same time.
- If we are not sure how the factors interact (epistemologic effect), we need to use some sort of factorial design (Like the Latin Hypercube Design);
- How to try and control factors – See Slide: Experimental Design (3) (Repetition of experiments, variation of the order, etc)
- Factor control: Expected factors (algorithm, problem), Possible factors (Software version, operating system, multi-tasking), unexpected factors (weather, color of the experimenter socks)
- Sometimes, we should blow a trumpet to tulips every morning for a month. Probably nothing would happen, but what if it did?

Proper Statistical Analysis!

- Follow the Appendix;
- The goal right now is not to get all the calculations right (we will do that in a future class), but to get the concept of statistical significance across.
- Definition of random variable – the result of an experiment is a random variable, if the experiment is not deterministic.
- The distribution of a random variable is the list of all possible values for all possible probabilities.
- Definition of Population and Sample, and how the mean of both can be different.

Statistical Significance Analysis Outline 2

- Show that as the sample changes, values such as the mean of the sample and the variance of the sample are also Random Variables;
- Random variables are not necessarily random;
- Therefore, if we repeat the experiment, we can have results that confirm or that deny or hypothesis, or that are very close. How can we know the validity of one particular result?
- In other words, we are asking how can we know if a good result is consistently good, in the face of random factors?

Statistical Analysis Outline 3

- Let us give a very simple example of how this is done.
- Again, the statistic formulas will be made explicit in the following classes. Today I just want you to understant the principle of statistical testing.
- We take a sample of a random variable (our experimental result), from this sample, we can calculate the means and the variance. If we assume that this random variable has a certain distribution, we can know the probability for many different values.
- For now, let's assume a Normal distribution. This is not always the best assumption, but it is a good one for our purposes (the assumed distribution of our random variable is a factor of the experiment!).

Statistical Analysis Outline 4

- Now that we have a distribution for our random variable, we define the hypothesis that we want to test, we do this by defining a null hypothesis and a alternate hypothesis.
- This hypothesis can be defined as a fixed value, or as a second random variable. And we can define it by relations of equality or inequality. Some Examples.
- This allows us to decide to "reject" the null hypothesis or "accept" the alternate hypothesis. Talk about Type I and Type II errors, remember that these values are a sample of a random variable!
- We can also think about minimum and maximum values for the null and alternate hypothesis, in order to minimize the probability of the errors.
- There are many formulas for this, but today I just want you to have a "intuition" for statistical testing. You will be sick of it by the end of the course.

Statistical Analysis Outline 5

- All the statistical analysis have assumptions about the data – we need to validate these assumptions.
- For example, assumption of normalty
- Not only that, but in some cases a statistically significant difference can be very small, if we have *N* very big. We need to calculate the scientific relevance of the result.
- We will talk about all of this in the next class.

Next Class

- Next Class we will talk about Statistics, how to use them, and R.
- Please bring your computers, with the “R” package installed in them (it is free!)
- If possible, also bring some text data from your research, maybe we will use it!