Experiment Design
Lecture 1: Introduction

Claus Aranha
caranha@cs.tsukuba.ac.jp

Department of Computer Science

April 13, 2015

Course Goal

Understand how to design, execute and analyze an experiment, in a Computer Science context.

Theoretical

Instrumental

Practical

Course Goal

Understand how to design, execute and analyze an experiment, in a Computer Science context.

Theoretical

What are the fundamentals of the Scientific method. How we recognize good/bad science, experiments, and scientific papers? Why are experiments necessary for our fields?

Instrumental

Practical

Course Goal

Understand how to design, execute and analyze an experiment, in a Computer Science context.

Theoretical

Instrumental

Statistical methods often used in an experimental context (hypothesis testing, significance, etc). How to analyse a data set. How to use the “R” tool to support statistical analysis of experiments.

Practical

Course Goal

Understand how to design, execute and analyze an experiment, in a Computer Science context.

Theoretical

Instrumental

Practical

How can I design an experiment for my own research? When I read other people's papers, can I spot and evaluate good/bad scientific practices? When considering a scientific hypothesis, or an engineering method, how to identify the assumptions of this research, and design experiments that test these assumptions or take them into account?

This is a participative course

Rount table philosophy

Context (the reality of your research) is very important in the design of experiments. I will very often ask you to give opinions based on your research experience!

Outline

- Class 1-2: Introduction to Science and Experimentalism. Common good and bad practices in Science. Reflect on the meaning of research and how this influence your work.
- Class 3: Basics of data handling: Understanding and visualizing data, simple statistics.
- Class 4: Simple R-tutorial
- 5/19 and 5/26: No class
- Class 5-9: Techniques for experimental design, testing and analysis. Parameter selection. Hypothesis testing. Result evaluation.
- Class 10: Report review.

Evaluation

Evaluation will be based on your ability to apply this classes ideas on your own research!

Main Report: Create an experimental design for your own research, using concepts introduced in this course. Define the factors, parameters, hypothesis, experiment and expected results. Use R to visually analyse your own data.

Report on Good/Bad Science: Write a review of two papers in your field of research (a good example, and a bad example). This review must be focused on the experimental design and analysis of the papers.

Just say NO to academic plagiarism

Useful Materials – Internet

- Course Webpage:
  `https://manaba.tsukuba.ac.jp/ct/course_397478`
  Manaba Code: 7739931
- Lecture Notes (Github):
  `https://github.com/caranha/ExperimentDesignLectureNotes`
- Lecture Notes by Felipe Campelo:
  `https://github.com/fcampelo/Design-and-Analysis-of-Experiments`
- R Programming Course:
  `https://class.coursera.org/rprog-011`

Useful Material – Books

- “Experimental Methods for the Analysis of Optimization Algorithms”, Thomas Bartz-Beielstein et al.
- “A Theoreticial Guide to the Experimental Analysis of Algorithms”, D.S. Johnson
- “Introductory Statistics with R”, Peter Dalgaard
- “Analysis of Categorical Data with R”, Christopher Bilder and Thomas M. Loughin

Let's get to know each other

- What is your research area?
- What kind of experiments do you normally do?
- What is hard in your research area?

What is Science?

Give me your opinions!

The Science Game - Question

- Imagine three numbers (triplets), and a rule behind them;
- Some triplets obey the rule, some don't;
- For example, the triplet $3, 6, 9$ obeys the rule!

Goal of the game:

- Find out the secret rule for the triplets;
- To find the rule, you can ask the judge (me), if any trio of numbers obey the rule or not;
- Feel free to discuss among yourselves!

The Science Game - Answer

Did you find the rule?

- What rule did you find?
- How did you arrive on that rule?
- How certain are you that this is the secret rule?

The Rule

The Science Game - Answer

Did you find the rule?

- What rule did you find?
- How did you arrive on that rule?
- How certain are you that this is the secret rule?

The Rule

Three Real Numbers in increasing order.

The Science Game - Insight

- Each question was “an experiment”;
- Did you use the experiments to confirm your idea, or to rule out other possibilities? (Confirmation Bias)
- What would be a good strategy for a similar game?
- When do you stop making questions?

Hans, the Smart Horse

Von Osten, the horse's trainer, displayed Hans amazing ability to do mathematical calculations by tapping his hoof.

Hans, the Smart Horse

Von Osten, the horse's trainer, displayed Hans amazing ability to do mathematical calculations by tapping his hoof.

Oskar Pfungst's Experiments

- Use people other than Van Osten to ask questions;

Hans, the Smart Horse

Von Osten, the horse's trainer, displayed Hans amazing ability to do mathematical calculations by tapping his hoof.

Oskar Pfungst's Experiments

- Use people other than Van Osten to ask questions;
- Put people in different places;

Hans, the Smart Horse

Von Osten, the horse's trainer, displayed Hans amazing ability to do mathematical calculations by tapping his hoof.

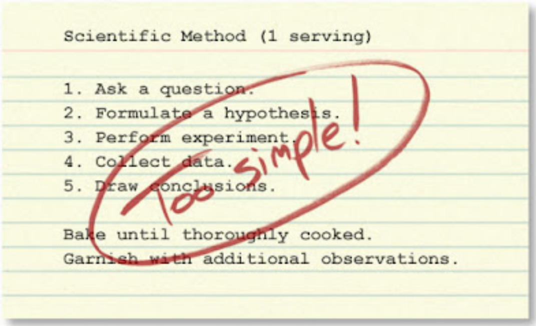
Oskar Pfungst's Experiments

- Use people other than Van Osten to ask questions;
- Put people in different places;
- Make people ask questions they didn't know the answer;

Common Misconceptions about Science

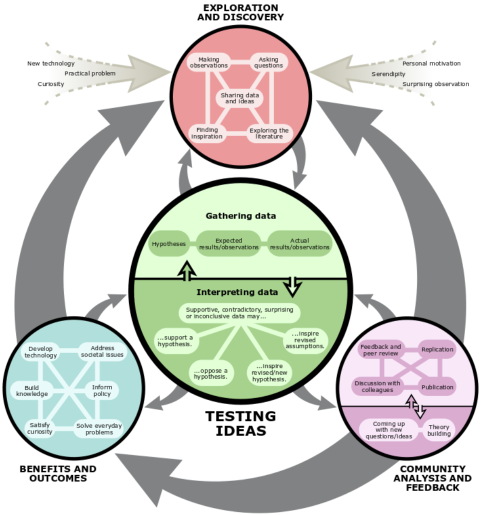
- Science is a collection of facts;
- Science is the creation of new gadgets;
- Scientific ideas are absolute and unchangeable;
- Scientific ideas are subject to change, thus unreliable;
- Observation gives answers directly to the scientists;
- Science proves stuff;
- Science can only disprove things;
- The scientist works to show his theory is right;
- Theory VS Law;

The Scientific Method



- A good start, but too simplified;
- The scientific method includes many iterations, interactions with multiple characters, etc;
- A set of principles, not a set of rules

The Scientific Method



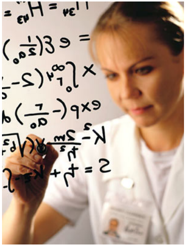
Important Considerations

- Science is about collecting data (evidence) to understand a phenomenon – but not hapzardly!
- “Science is about being Wrong” – strive to look for evidence that the idea you hold is incorrect;

The role of experimentation

- In this sense, the experiment is one of the main ways of collecting data;
- But before we get to the experiment, we will think about some questions you may have;

Maybe Science is not for me...



I don't really need experiments. All of my research involves well behaved, deterministic algorithms. Mathematical proofs are enough for me.

I don't really need experiments. My research involves creating a completely new service/device. If people seem excited about using this, then I'm doing it right.

Uncertainties Everywhere

When it comes to gathering data and experimental rigour, Researchers in Computer Science are still years behind other disciplines. A large disconnect still exists between the "theorists" and the "empirists" in CS.

The Theorist's lament

Even if the algorithm is mathematically perfect, implementation details may affect the performance of the algorithm, or sometimes even the behavior. Many times, it is worth measurings the interaction betwee the system we are interested in, and other nsources.

The Empirist's folly

Ad-hoc experiments hold many valuable lessons, but often they can't be used to get quantitatie information. Without care, uncertainties in the experiment may weaken the result.

Sources of Uncertainty

- Physical Systems;
- Implementation Details;
- Random Number Generator;
- People in the experiment;

Physical Systems

Transporting an algorithm from the blackboard to a production machine (or to a physical device) will include uncertainties from the interaction with the real world. (e.g.: Machine load, errors in robotic torque, etc)

Implementation Details

Performance of an algorithm might vary greatly based:

- Programming techniques (data structures, design choices);
- Library used;
- Operating System;
- etc.

Random Number Generation

It is very difficult to prove an algorithmic behavior if part of it is defined by randomly generated numbers.

Implementation of the RNG library can have a huge effect on algorithm behavior.

Humans

Including humans in your research (eg. Interfaces), is a huge source of uncertainty regarding the results.

You have to take into account not only how the humans will act, but also what are the characteristics of the particular group that you are using (don't do all your experiments on college students).

Experiment Design

By carefully designing the experimental methods of our research, we can account for these uncertainties in our results.

- Understand the influence of the uncertainty in the final results; or
- Take into account the state of the uncertainty as a condition for the experiment;

Lack of an Experiment Design

A common misunderstanding

Science is about creating cool gadgets and running many experiments.

Lack of an Experiment Design

A common misunderstanding

Science is about creating cool gadgets and running many experiments.

- Experiments that only show the system works;
  (of course it works, but WHY? And why is it good?)

Lack of an Experiment Design

A common misunderstanding

Science is about creating cool gadgets and running many experiments.

- Experiments that only show the system works;
  (of course it works, but WHY? And why is it good?)
- Reinventing the wheel;
  (Lack of comparisons and inspiration from other methods)

Lack of an Experiment Design

A common misunderstanding

Science is about creating cool gadgets and running many experiments.

- Experiments that only show the system works;
  (of course it works, but WHY? And why is it good?)
- Reinventing the wheel;
  (Lack of comparisons and inspiration from other methods)
- No knowledge is actually acquired;
  (unreplicable papers, excess of data)

Experiment Design to the Rescue

How can we know if the effect we are observing is a result of the research being tested, or if it is an artifact from the uncertainties.

Experiment Design to the Rescue

How can we know if the effect we are observing is a result of the research being tested, or if it is an artifact from the uncertainties.

Pshysical Uncertainties

Experiment Design to the Rescue

How can we know if the effect we are observing is a result of the research being tested, or if it is an artifact from the uncertainties.

Implementation

Experiment Design to the Rescue

How can we know if the effect we are observing is a result of the research being tested, or if it is an artifact from the uncertainties.

Random Numbers

Experiment Design to the Rescue

How can we know if the effect we are observing is a result of the research being tested, or if it is an artifact from the uncertainties.

Human Factor

Goals for an experiment

The main characteristics that we look for in a good experiment are:

Clarity: Does the experiment really show the kind of information that we are interested in? Can it be used to confirm or reject our hypothesis? Does it show any other interesting data?

Replicability: Can other people repeat my experiment 5, 10 years in the future? Are all algorithms, parameters and data well defined? Am I sharing my source code?

Control: How many “moving wheels” and “conditions” does my experiment has? Do I know how general the results obtained are?

On experiments

"However, experiments require a lot of work, so the reader may be warned: performing a good experiment is as demanding as proving a new theorem."

– Hans-Paul Schwefel

Experiment Design: Control

- Decide what is the quantity to be examined;
- Control the values of all other quantities;
- Define the range and index for the variation;

There are principles and rules for defining these designs, which we will watch carefully in the next class.

Experiment Design: Replicability

- If other people follow our steps, can they reach the same results as we did?
- Extremely important not only for validation, but also for other people to use your research;
- Make data, parameters and source code available;
- Take special care with small tricks and hacks that you may have added during your work;

Experiment Design: Clarity

- Besides clarity regarding the replicability of your research, clarity regarding the scope and achievements of your work are important;
- In which cases is your research valid?
- What are your assumptions?

What are the assumptions made?

A very narrow claim with clear assumptions is better than a wide claim that does not state its assumptions.

Experiments: The Ugly

News link: “Two thirds of drug studies failed to replicate”

- Not all data available;
- Not all experimental parameters available;
- Some details of the experimental method were not listed (hacks)

(Specially bad in the computer sciences)

Bad Practices

Careless Experimentation

"To Consult the statician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." – Sir Ronald Fischer

- People are well meaning (I hope!), but sloppy research still gets done.

Bad Practices

Careless Experimentation

"To Consult the statician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." – Sir Ronald Fischer

- People are well meaning (I hope!), but sloppy research still gets done.
- Lack of training, Lack of awareness, Publication bias;

Bad Practices

Careless Experimentation

"To Consult the statician after an experiment is finished is often merely to ask him to conduct a post mortem examination. He can perhaps say what the experiment died of." – Sir Ronald Fischer

- People are well meaning (I hope!), but sloppy research still gets done.
- Lack of training, Lack of awareness, Publication bias;
- Correct Goal: Finding out the truth;
- Common Goal: Finding out that the algorithm is publishable;

Careless Experimentation

"The presented results represent the average from 30 runs of each algorithm"

Careless Experimentation

"The presented results represent the average from 30 runs of each algorithm"

What "average of 30 runs" does not say

Proposed Algorithm

- Careful implementation and error checking;
- Exhaustive search of appropriate parameters;
- Many Experimental runs;

Comparison Algorithm

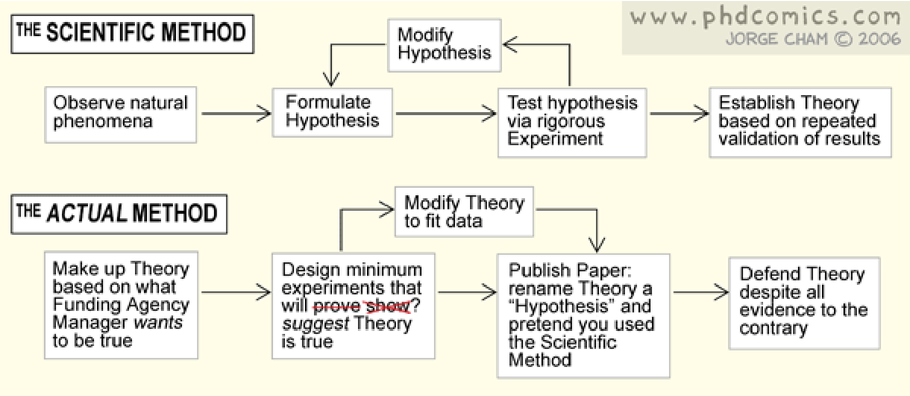
- Careless implementation: Good enough if it runs;
- Parameters found in the literature;
- Single experimental run;

Careless Experimentation

Other Common Problems

- No clear definition of the Goal of the experiment (confusion among hypothesis, theory, explanation, conjecture)
- Cherry picking of data/results (This data that does not work falls outside the scope of this paper).
- Personal Biases (getting experimentation backwards: trying to show that one method **is** better, instead of investigating **if** the method is better).

Careless Experimentation



Other common mistakes

Premature conclusions

- Stop as soon as a successful result is achieved;
- No random repetitions, no stress tests;
- Even some negative results can be informative!

Conclusion

Rigorous scientific methods are important for computer science, in order to isolate uncertainties and quantify our discoveries.

For the Next Class

- Each of you should prepare a new short introduction to your disciplines/research topic;
- Take into account the topics discussed today;
- How would you apply this knowledge to your existing research;

One final message

It is important to be literate in the scientific method, not only for your own research's sake. We are also agents of change in the population and, as such, we need to be aware of good and bad science, and able to point the difference to the society.

Acknowledgements

- Felipe Campelo, for many comments and insights on the contents of this course;
- "The Science Game" inspired from `http://hpmor.com`;
- "Common Misconceptions About Science" `http://undsci.berkeley.edu/teaching/misconceptions.php`

Contact Me!

E-mail: caranha@cs.tsukuba.ac.jp
Homepage: `http://conclave.cs.tsukuba.ac.jp`
Twitter: @caranha