Physics 509C: Theory of Measurement

Scott Oser
Lecture #1

TYXH (Greek goddess of chance)

Today's Outline

- Course content/syllabus
- Course mechanics
- Computing/software aspects of the course
- What is probability?
- Frequentist vs. Bayesian interpretations
- Bayes theorem
- Probability distribution functions (PDFs)

What is this course about?

Primarily this course is about making inferences from uncertain data without embarrassing yourself.

Few physicists or astronomers entirely succeed at this.

You will learn:

- What error bars really mean, how to treat them, and how to recognize when they are being abused
- How to reason probabilistically about data
- How to propagate uncertainties correctly
- How to estimate model parameters and decide between hypotheses
- How to do the relevant computation and techniques
- Practical approach to probability and statistics

An extremely broad topic---we will have to pick and choose topics.

Tentative Course Syllabus

In my happiest dreams we will cover all of the following ...

- Interpretation of probability
- Basic descriptive statistics
- Common probability distributions
- Monte Carlo methods
- Bayesian analysis
- Methods of error propagation
- How to handle systematic uncertainties
- Parameter estimation
- Hypothesis testing and statistical significance
- Confidence intervals
- “Blind analysis”
- Methods of multivariate analysis
- Non-parametric tests
- “Robust” statistics
- Deconvolution & unfolding

Textbook(s)

Primary texts (I will mostly draw on these for lectures, and recommend that you purchase them):

- *Statistics: A Guide to the Use of Statistical Methods in the Physical Sciences*, by R.J. Barlow. Strong frequentist introduction.
- *Bayesian Logical Data Analysis for the Physical Sciences,* by Phil Gregory. One of the best in-depth treatments of Bayesian techniques available.

Other texts we will draw upon:

- *Numerical Recipes*, by William H. Press *et al.* Any serious scientist should own this book. Text freely available online.
- *Statistical Data Analysis,* by Glen Cowan. A model of conciseness and clarity.
- *Practical Statistics for Astronomers,* by J.V. Wall and C.R. Jenkins. Many astro-specific examples.
- *Probability and Statistics,* by Morris H. DeGroot. For the mathematically inclined.

Course Format

Lectures: TuTh 14:00-15:15

TA: Aditi Pradeep

Webpages: http://www.phas.ubc.ca/~oser/p509/
Plus we will use Canvas for submitting assignments

Problem Sets: I will assign ~5 substantial problem sets spaced throughout the course. All will involve some computational component in additional to analytic analysis and/or essay/conceptual questions. (The password to access them is posted on Canvas!)

In-class midterm: October 17

Take-home final: will include some computational component.

Office Hours: Mondays 12-1pm, or by appointment

In person and online components

All lectures are in person.

Office hours will be conducted over Zoom. Connection information is posted on the course's Canvas page.

All HW, and the final exam, must be submitted by uploading them to Canvas. Please submit a single PDF or Word document. Avoid submitting multiples files. If you prefer to work your results on pen and paper, please scan or photograph the pages and assemble the entire assignment (including any code you're submitting) into a single file.

Some people have recommended these apps for help with scanning and making a PDF file for submission. It basically lets you use the camera function of your phone like a scanner:

https://www.camscanner.com/

https://apps.apple.com/us/app/genius-scan-pdf-scanner/id377672876

https://play.google.com/store/apps/details?id=com.thegrizzlylabs.geniusscan.free&hl=en_US

Or you can photograph the individual pages and paste them into a Word document (one image per page), and submit the Word document.

“Discussion Days”

The beauty of this course is that it almost certainly is relevant to your own research interests.

Accordingly I've introduced two “discussion days”. For these classes I will ask you to submit questions for discussion, and we'll pick some and brainstorm together about them. For example, if you have a relevant problem in your research you want advice about, you can share it with the class and we can discuss it together. We can also discuss HW problems.

This will work only if people come prepared to discuss things.

First full discussion day: November 2.

Computing

Statistics and data analysis are naturally computationally intensive. The class of problems you can solve by hand is very, very small!

You will need to write your own programs or adapt existing tools (ex. Matlab, Mathematica) to complete the HW assignments. I will assume that you know some high-level programming language (e.g. C, C++, python, FORTRAN, Java), or that you are willing and ready to learn one as you go. I don't care which you use.

I speak C and FORTRAN well, and some C++ in addition. I will try my best to answer general questions about what approach you might take, but I will not debug your code for you. Your best resources for computing help are your fellow students.

If this prospect seems too daunting to you, your time might be better spent taking Physics 410 (Computational Physics).

Plotting packages

Plotting package: the HW will often request that you produce plots of a distribution or function. Plots should be legible and clearly labelled with a title and axis labels.

I don't care what plotting package you use—-whatever you feel comfortable with. Some suggestions if you don't already have a favourite:

- Gnuplot (www.gnuplot.info): free, widely used, with some basic fitting capability (probably not sufficient to do all HW)
- ROOT (root.cern.ch): free, CERN-supported, based on C++ syntax. Designed with particle physics in mind but not application-specific in itself. Very powerful library of statistics and numerical routines built in. Integrated plotting and analysis/coding package.
- Matplotlib: visualization library for Python
- Mathematica: I don't use it myself, but be my guest if you already know it.

Libraries of numerical algorithms

Don't re-invent the wheel. If you need to invert a matrix, minimize a non-linear function, or do some complicated mathematical operation, use pre-existing code! Some suggestions:

- Numerical Recipes: proprietary, but “industry standard”. Widely used. Text itself is a great reference even if you don't use the code itself. Available in C, FORTRAN, C++. (http://www.nr.com)
- NumPy: useful if you’re a python person (http://numpy.org)
- GSL (GNU Scientific Library): free C/C++ code providing equivalents for most Numerical Recipes routines in an “open source” model. (http:www.gnu.org/software/gsl)
- ROOT math libraries (MINUIT minimizer, linear algebra, random number routines, etc.): if you use ROOT for plotting, you can write your analysis program within ROOT as well, since ROOT is basically a fancy C++ interpreter with plotting capacity.
- Mathematica: proprietary, expensive, but powerful. I cannot guarantee that all HW problems can be solved with it, but many can.

Typical computational problems we'll face

- Minimizing functions of several variables and describing the shape of the function around its minimum--contours, curvature terms (second derivatives), etc.
- Random number generation
- Solving systems of linear equations
- Matrix manipulation, including matrix inversion
- Numerical integration, including multi-dimensional integration
- Sorting

Although I will not teach you how to program, we will discuss relevant computational methods when applicable.

HW guidelines

- You are welcome, even encouraged, to discuss the assignments among yourselves and share insights. But in the end you must complete the work yourself. *Do not hand it someone else's work!*
- Each assignment will require some programming, and you should submit the code or equivalent documentation as part of the assignment. I will *not* run the code myself, but we will look at it while grading the assignment. Please liberally comment your code (for your benefit as much as ours) so that it is readable and understandable! It's not enough to just submit your answer.
- First assignment will require very basic programming, just to get your feet wet, and is due on Sept 22.

Reminder about UBC policy on academic misconduct:

http://www.calendar.ubc.ca/Vancouver/index.cfm?tree=3,54,111,959

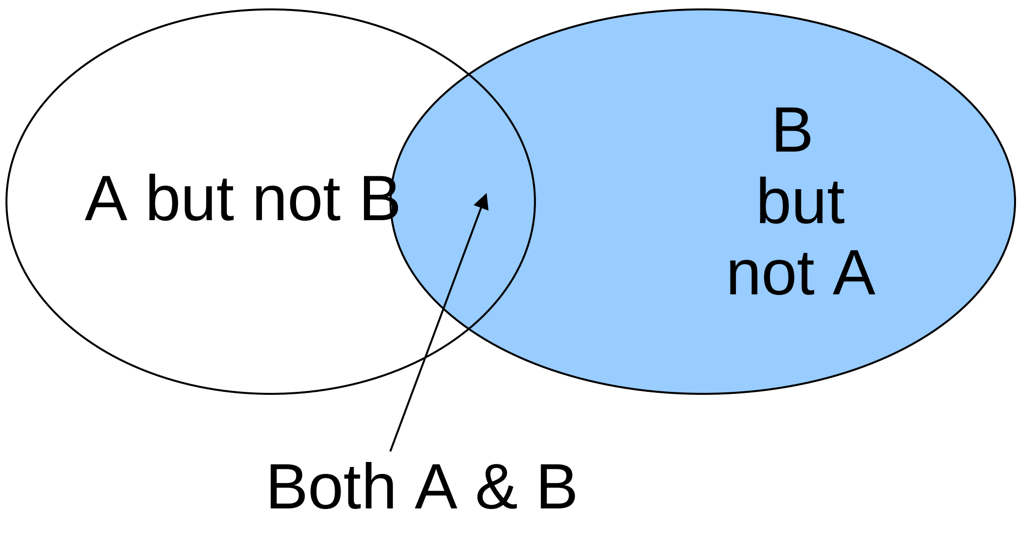
What is probability?

Kolomogorov's axioms:

1) The probability of an event E is a real number P(E)≥0.
2) If two events E_1 and E_2 are mutually exclusive, then
   $P(E_1 \text{ or } E_2) = P(E_1) + P(E_2)$
3) Summing over all possible mutually exclusive outcomes, we get

$$\sum P(E_i) = 1$$

All of probability follows from these axioms ... for example:



$$P(A \text{ or } B) = P(A) + P(B) - P(A \wedge B)$$

But what does P mean?

Interpretations of probability

There are multiple, sometimes mutually exclusive, ways to interpret probability. *WHICH DO YOU BELIEVE?*

1) Probability is a statement about frequency. If you repeat a measure 1000 times and get the same outcome 200 times, the probability of that outcome is 0.2.

2) Probability is an intrinsic property of objects with its own objective existence. QM predicts probabilities---they somehow are intrinsic properties of systems.

3) Probability is a statement about our knowledge, and so is subjective. While I say the probability of rain tomorrow is 1/3, you may have reason to believe otherwise and assign a different probability.

Problems with the frequency interpretation

1) We naturally want to talk about the probability of events that are not repeatable even in principle. Tomorrow only happens once---can we meaningfully talk about it? Maybe we want to talk about the probability of some cosmological parameter, but we only have one universe! A strict interpretation of probability as frequency says that we cannot use the concept of probability in this way.

2) Probability depends on the choice of ensemble you compare to. The probability of someone in a crowd of people being a physicist depends on whether you are talking about a crowd at a hockey game, a crowd at a university club, or a crowd at a CAP meeting.

In spite of these conceptual problems, the “frequentist interpretation” is the most usual interpretation used in science.

Problems with the "intrinsic propensity" interpretation

1) To assign objective reality to probabilities is certainly engaging in metaphysics and not just physics. We never can see or even measure a probability directly---all we can measure are frequencies of outcomes.

2) Sometimes we clearly are using probability in a subjective sense! I hide a coin in one hand. You say that there is a 50% probability that it's in my right hand, and this strikes most of us as a meaningful statement. But of course I know with 100% probability which hand the coin is in!

The “subjective” interpretation

This goes most commonly by the name “Bayesian statistics”. In this view probability is a way of quantifying our knowledge of a situation. P(E)=1 means that it is 100% certain that E is the case. Our estimation of P depends on how much information we have available, and is subject to revision.

The Bayesian interpretation is the cleanest conceptually, and actually is the oldest interpretation. Although it is gaining in popularity in recent years, it's still considered “heretical”. The main objections are:

1) “Subjective” estimates have no place in science, which is supposed to be an objective subject.
2) It is not always obvious how to quantify the prior state of our knowledge upon which we base our probability estimate.

Purely anecdotal personal observation: the most common reason for scientists not to be Bayesian is because they think that most other scientists aren't.

Frequentist vs. Bayesian: does it matter?

You might hope that such issues would be of philosophical interest only, and as relevant to science as the hundreds of interpretations of QM.

Unfortunately it DOES matter. The interpretive framework determines which questions we ask, how we try to answer them, and what conclusions we draw.

This course will attempt to make you “bilingual”, comfortable in both schools of thought. (Hence the two recommended textbooks.)

In general we will spend more time on frequentist methods, which are most commonly encountered, but will cover Bayesian techniques thoroughly enough that you can use them when appropriate or desired. In many cases the Bayesian approach is simpler to understand---we will even resort to justifying some frequentist formulas by looking at Bayesian approaches.

However, be careful to be clear what interpretation you are using and to avoid inconsistency. More on this later.

Random Variables

Consider the outcome of a coin flip.

Use the symbol "b" to represent the observed outcome of the coin flip. Either b=1 ("heads") or b=0 ("tails"). Note that b has a known value---it is not considered to be random.

Let B represent the possible outcome of the next coin flip. B is unknown, and is called a "random variable".

Random variables are used to represent data NOT YET OBSERVED.

Although we don't know what the value of B will be, there is other information about B that we may know, such as the probability that B will equal 1.

In frequentist language:

$$P(B=1)=\lim \frac{\text{number of occurrences of } b=1 \text{ in } n \text{ trials}}{n}$$

Frequentist vs. Bayesian Comparison

Bayesian Approach

- "The probability of the particle's mass being between 1020 and 1040 MeV is 98%."
- Considers the data to be known and fixed, and calculates probabilities of hypotheses or parameters.
- Requires *a priori* estimation of the model's likelihood, naturally incorporating prior knowledge.
- Well-defined, automated "recipe" for handling almost all problems.
- Requires a model of the data.

Frequentist Approach

- "If the true value of the particle's mass is 1030 MeV, then if we repeated the experiment 100 times only twice would we get a measurement smaller than 1020 or bigger than 1040."
- Considers the model parameters to be fixed (but unknown), and calculates the probability of the data given those parameters.
- Uses "random variables" to model the outcome of unobserved data.
- Many "ad hoc" approaches required depending on question being asked. Not all consistent!
- Requires a model of the data.

Basic mathematics of probability

1) P(A or B) = P(A) + P(B) – P(A & B)

2) Conditional probability: P(A & B) = P(B) P(A|B).
Read "the probability of B times the probability of A given B".

3) A special case of conditional probability: if A and B are *independent* of each other (nothing connects them), then

P(A & B) = P(A) P(B)

Bayes' Theorem

$$P(H|D,I) = \frac{P(H|I)\,P(D|H,I)}{P(D|I)}$$

This just follows from laws of conditional probability---even frequentists agree, but they give it a different interpretation.

H = a hypothesis (e.g. "this material is a superconductor")
I = prior knowledge or data about H
D = the data

P(H|I) = the "prior probability" for H

P(D|H,I) = the probability of measuring D, given H and I. Also called the "likelihood"

P(D|I) = a normalizing constant: the probability that D would have happened anyway, whether or not H is true.

Note: you can only calculate P(D|I) if you have a "hypothesis space" you're comparing to. A hypothesis is only "true" relative to some set of alternatives.

An example ...

http://math.ucsd.edu/~crypto/Monty/monty.html

Discussion: Triple Screen Test

The incidence of Down's syndrome is 1 in 1000 births. A triple screen test is a test performed on the mother's blood during pregnancy to diagnose Down's. The manufacturer of the test claims an 85% detection rate and a 1% false positive rate.

You (or your partner) test positive. What are the chances that your child actually has Down's?

Discussion: Triple Screen Test

The incidence of Down's syndrome is 1 in 1000 births. A triple screen test is a test performed on the mother's blood during pregnancy to diagnose Down's. The manufacturer of the test claims an 85% detection rate and a 1% false positive rate.

You (or your partner) test positive. What are the chances that your child actually has Down's?

Consider 100,000 mothers being tested. Of these, 100,000/1000=100 actually carry a Down's child, while 99,900 don't. For these groups:

85 are correctly diagnosed with Down's.
15 are missed by the test
999 are incorrectly diagnosed with Down's
98901 are correctly declared to be free of Down's

Fraction of fetuses testing positive who really have the disorder:
85/(85+999) = 7.8%

Discussion: Triple Screen Test

The incidence of Down's syndrome is 1 in 1000 births. A triple screen test is a test performed on the mother's blood during pregnancy to diagnose Down's. The manufacturer of the test claims an 85% detection rate and a 1% false positive rate.

You (or your partner) test positive. What are the chances that your child actually has Down's? We calculated 7.8%.

Some questions for thought:

- Are all mothers equally likely to have a triple screen test performed? If not, how would this affect the result?
- You read that 90% of pregnancies in which a Down's diagnosis is made are aborted. How does this change the result?
- What is the relevant “prior” in any individual case?

Probability Distribution Function

Discrete distribution:

P(H) = probability of H being true

Ex. H=”rolling two dice gives a total of 7”

Continuous distribution:

P(x) dx = probability that x lies in the range (x, x+dx)

Ex. probability of mean of N measurements being between 5.00 and 5.01

NORMALIZATION CONDITION:

$$\sum P(H_i) = 1 \qquad \text{or} \qquad \int dx\, P(x) = 1$$

EXTRA MATERIAL

Probability as a generalization of Aristotlean logic

Compare:
1) If A is true, then B is true.
2) B is false.
3) Therefore A is false

with:
1) If A is true, then B is true
2) B is probably not true
3) Having learned that B is probably not true, I am less convinced than A is true.

The “subjective” interpretation of probability can be considered to be an attempt to generalize from deductive logic.

Desiderata of inductive reasoning

Here's how to “derive” the rules of probability. Demand that:

1) Degrees of plausibility are represented by real numbers
2) “Common sense”: as data supporting a hypothesis is accumulated, the plausibility of the hypothesis should increase continuously and monotonically.
3) Consistency: if there are two valid ways to calculate a probability, both methods should give the same answer!

An amazing result: if you try to construct a system of assigning degrees of plausibility to statements according to these requirements, the only unique way to do so results in the regular rules of probability!

Conclusion: probability is the unique inductive generalization of “Boolean algebra”

Objective vs. Subjective

Ed Jaynes imagined a robot programmed to use a Bayesian interpretation of probability:

“Anyone who has the same information, but comes to a different conclusion than our robot, is necessarily violating one of those desiderata. While nobody has the authority to forbid such violations, it appears to us that a rational person, should he discover that he was violating one of them, would wish to revise his thinking ...”

In other words, Bayesian probability estimates are still objective provided that any two observers starting with the same information and looking at the same data will assign the same probability estimates to the result.