Instructor: Dick Furnstahl

8/21/19

Physics 8805: Learning from Data: Bayesian methods and (some) machine learning; Lecture 1

Handout: Hardcopy of simple_sum_product_rule.ipynb

Before class: Set up projector with laptop showing Carmen/Canvas 8805 pages (Syllabus and Modules) in Student View.

- On board: Welcome to Physics 8805: Learning from Data. Please bring up the Carmen page on a laptop (local or your own). If local laptop, set default browser to Google Chrome [in search box, type: Default Apps, from this page Web Browser → Chrome]
- If early try Jupyter notebooks under Modules → Python and Jupyter Notebook Resources → Binder links for Jupyter Notebooks. If you have Anaconda installed try downloading notebooks from 0. and 1. modules.

Overview of today:

- Quick review of syllabus and Canvas modules
  ⇒ scope of course and resources
- Put some Bayesian concepts on the table
  ⇒ Philosophy based on learning to swim by throwing you in the water.
  - absorb details as we go through notebooks
  - spiral method
  - can be confusing and sometimes frustrating. But it works and is good training for real-world situations.

⇒ Ask questions. Question authority. Experiment. Verify everything.

Survey summary:

- A couple of stats experts, but mostly limited prior knowledge
- Very few with Bayesian statistics experience and over half with no sampling (MCMC) or Gaussian processes ⇒ I won't assume anything.
- Some programming experience and at least a little Python but less Jupyter. Will build this up as we go. Help each other!

8/21/19

Syllabus overview:

- We have the room before 11am but another class at 12:40pm.

My story: Nuclear theorist doing effective field theory, which means building description's of nuclear properties as an expansion: "next to leading order"

Hamiltonian $\Rightarrow$ $H_{LO} + H_{NLO} + H_{N^2LO} + \ldots$ $\leftarrow$ parameters to fit

Energy: $E_{nucleus} = E_{LO} + E_{NLO} + E_{N^2LO} + \ldots$ $\quad E_{LO} > E_{NLO} > E_{N^2LO} > \ldots$

- If we truncate the theory at some order, say N^3LO:

sample questions

- What are the uncertainties in the parameters because the theory is incomplete?
- How do we optimally take advantage of data we fit to if theory works better at low energy?
- What is the truncation error for $E_{nucleus}$? This is a systematic error, not a random one (like counting error)

Answer to all: Use Bayesian statistics! Also: how to distinguish between alternative theories?

- These issues are not restricted to nuclear physics $\Rightarrow$ common problems (and many more).

- Some fields like astronomy and cosmology went heavily Bayesian long ago. Others like hadron physics are skeptical (Jordan Melendez reports skeptical community at HADRON 2019 in Guilin.)

- Nuclear physics is in transition to Bayesian $\Rightarrow$ education needed $\Rightarrow$ TALENT course $\Rightarrow$ This course.

8/21/19

Prerequisites: Just assuming undergrad lab statistics and first-year grad. physics.

Topics: Comment on coverage.
- Already a lot to get through, but you can lobby for others.
- Machine learning will be limited but could be ramped up.

Learning outcomes: Quick review of what you should take away from the course. Can't be exhaustive!

Textbooks:
- No required text.
- Highlight Sivia; excerpts will be posted.
- Gregory and Trotta are good (examples from Gregory) for physicists.
- BDA3 is standard but thick and focus is social sciences.
- More machine learning refs. later.

Grading: Jupyter notebooks, mini projects, final project.
- Should not be onerous workload.
- But you can supplement and do more!
- Final project integrated in your research. (Product is a notebook like those for mini-projects (Could be reproducing figures in a paper or making a new application of a topic from class.)
- Plus, check, minus system

8/27/19

Course Module overview

- Printing in Smith 1094
- Look at Binder links → try one out (Anaconda ~~test~~ notebook) getting started
  - Server disconnects if inactive for ~15 minutes
- Cheatsheets → use Google more often but good for review.
- Instructions for Anaconda

Resources

How many have installed? →

0. Getting started
   - look at a few features, but mostly leave for you to explore and ask questions.
   - see Take-away points in notes.
   - You can use Anaconda Navigator or work from a terminal window.

I. Basics of Bayesian statistics
   - Please read the first chapters of Sivia and Gregory.
   - We'll start today with Exploring_pdfs.ipynb and simple_sum_product_rule.ipynb.

8/21/19

Take-away points from TALENT_Jupyter_Python_intro_01.ipynb

(1) Know how to switch between Code and Markdown cells and how to run them (shift-return or Run button)

(2) basics of Python evaluation and printing
- string concatenation
- exponentiation with **
- use of fstrings

(3) Importing numpy and basic functions (sqrt, exp, sin, ...)
- numpy arrays using arange(min, max, step)
- functions of arrays

(4) Getting help via Google: Stack Overflow and manuals

(5) Defining functions in Python
- def my_function(x): ← semicolon and then indentation
- positional vs. keyword arguments and defaults (more later)
- shift + Tab + Tab to reveal definitions

(6) Plotting with Matplotlib
- Standard sequence for plot: data → make figure → add subplots → make plot
- Dressing up and saving a plot
- Names for the figure and axis objects are our choice
- %matplotlib inline to generate inline plots in Jupyter notebooks

8/21/19

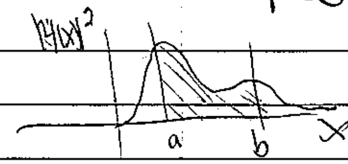
Inference: Pass 1

- deductive inference: cause → effect
- inference to best explanation: effect → cause

Scientists need ways to:
a) quantify the strength of inductive inferences
b) update that quantification as new data is acquired

Bayesian: Do this with pdfs ⇒ probability distribution functions
- Can be discrete or continuous.
  - We will mostly do continuous ⇒ probability density function is what we usually mean by pdf.
- To a Bayesian, everything is a pdf!

—3 Let's use physics examples: normalized wave functions squared.
—2 · discrete example: spin wave function. $p_{up} + p_{down} = 1$
$|\psi_1|^2+|\psi_2|^2+|\psi_3|^2 = 1$ · continuous example: one-dimension particle in coordinate space



Probability density at x: $|\psi(x)|^2 \Rightarrow p(x)$
- Remember that this has units
- The probability (dimensionless) of finding x in $a \le x \le b$ is

$$\text{prob}(a \le x \le b) = \int_a^b |\psi(x)|^2 \, dx \quad \Leftarrow \text{here } p(x) \sim \frac{1}{[L]} \text{ units}$$

Be careful about probability vs. probability density.

- multidimensional normalized pdfs: wavefunction squared for particle 1 at x_1 and particle 2 at x_2:

$$|\Psi(x_1, x_2)|^2 \Rightarrow p(x_1, x_2) = p(\vec{x}) \quad \text{with } \vec{x} = \{x_1, x_2\}.$$

8/21/19

Alternative notation in literature: $p(\tilde{x}) = P(\tilde{x}) = pr(\tilde{x}) = prob(\tilde{x}) = \ldots$

- Vocabulary and definitions:
  - $p(x_1, x_2)$ is the joint probability density of x_1 and x_2

class: (tell your neighbor)
  - What is probability to find particle 1 at x_1 and particle 2 anywhere? $\Rightarrow \int |\psi(x_1, x_2)|^2 \, dx_2$ integrated over domain of x_2 (e.g. $-\infty$ to ∞)

  - General: marginal probability density of x_1 is $p(x_1) = \int p(x_1, x_2) \, dx_2$.

  - "Marginalizing" = "integrating out" (eliminates "nuisance parameters").

- In Bayesian statistics there are pdfs (or pmfs if discrete) for ("mass")
  - fit parameters – like slope and intercept
  - experimental and theoretical uncertainties
  - measured quantities
  - hyperparameters (more later!)
  - events ("Will it rain tomorrow")
  - and much more.

Class! - What is the pdf $p(x)$ if we know definitely that $x = x_0$ (fixed)?
Answer: $p(x) = \delta(x - x_0)$ [note that it is normalized]

- Questions?

8/21/19

you will do this

First look at visualizing pdfs ⇒ come back and play with notebook $\boxed{\text{Exploring\_pdfs.ipynb}}$

Points of interest:

- %matplotlib inline

- importing packages: scipy.stats, numpy, matplotlib. Convenient abbreviations.
- corner is not included in Anaconda ⇒ use package manager conda. Google "conda corner" to find the command needed.

⇒ look for Corner :: Anaconda Cloud ⇒ conda install -c astropy corner

- scipy.stats ⇒ look at manual page.
- come back and look at definitions

- Look at examples: not everything is a Gaussian distribution!
  - You will look at Student t pdf on your own. (Trivia: Student was the pen name of the Head Brewer of Guiness – pioneer of small sample experimental design. Real name was William Sealy Gossett.)

- Look at projected posterior plots using the corner package.
  - ✗ With neighbor: what do you learn from plots?
  - Note that these are samples from the pdf. We will have much to say about sampling.

- 1d pdfs: note the fluctuations, larger for smaller numbers of samples.

Many follow-ups, but let's put some other Bayesian notions on the table first.

8/21/19

Manipulating pdfs: Bayesian rules of probability as principles of logic.

- You will show these rules are consistent with standard probabilities based on frequencies in simple_sum_product_rule.ipynb (also available as handout)

- Notation: $p(A|B)$ = "probability of A given B is true"
  $\Rightarrow$ conditional probability.
  For a Bayesian, A and B could stand for almost anything.
  Give examples: $p(\text{"below zero temperature"} | \text{"it is January"})$
  $$p(x|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-\mu)^2/\sigma^2}$$

- NOTE $\underset{\text{conditional}}{p(A|B)} \neq \underset{\text{joint}}{p(A,B)}$

- In the examples here, $p(x|I)$ is probability or pdf of x being true given information I.

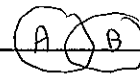
1. Sum rule: If set $\{x_i\}$ is exhaustive and exclusive

$$\underset{\text{(discrete like spins)}}{\sum_i p(x_i|I) = 1} \longrightarrow \underset{\text{(continuous)}}{\int dx\, p(x|I) = 1}$$ sum of probabilities is one.

- exhaustive, exclusive $\Rightarrow$ cf. complete, orthonormal
- implies marginalization (cf. inserting complete set of states or integrating out variables – but be careful, the analogy breaks down)

$$p(x|I) = \sum_j p(x, y_j|I) \rightarrow p(x|I) = \int dy\, p(x,y|I)$$

- We will use marginalization a lot!

cf.  $p(\underset{\uparrow\ \text{union}}{A \cup B}) = p(A) + p(B) - p(\underset{\uparrow\ \text{intersection}}{A \cap B})$ not exclusive

8/20/19

2. Product rule: expanding a joint probability of x and y

$$p(x,y|I) = p(x|y,I)\,p(y,I) = p(y|x,I)\,p(x,I)$$

- Note the symmetry.
- If x and y are <u>mutually independent</u>: $p(x|y,I) = p(x|I)$, then

$$p(x,y|I) = p(x|I) * p(y|I)$$

- Rearranging the 2nd inequality above yields Bayes' Rule (or Theorem):

$$p(x|y,I) = \frac{p(y|x,I)\,p(x|I)}{p(y|I)}$$

- Tells us how to reverse conditional $p(x|y) \rightarrow p(y|x)$.
- Tells us how to <u>update</u> expectations (more to come!)

Proof of 1., 2. by Richard Cox: "The Algebra of Probable Inference."

- Your task: complete the notebook simple_sum_product_rule.ipynb
  - Fill in table based on knowledge/intuition of probabilities as frequencies (these are <u>estimates</u> of population probabilities).
  - Apply sum and product rules as directed.
- Work together and check answers with each other.
- Ask questions!
- We'll recap when done.