9/13/19

Physics 8805: Learning from Data: Lecture 7

On board:

Notebooks for today:

- Metropolis_Poisson_example.ipynb
- MCMC-random-walk-and-sampling.ipynb

comment on negative y_0

- Assignment 1 is due via email to me by midnight tonight. If you are having problems, office hours 3-5 pm in M2048 in PRB this afternoon, or email me. [Goal: everyone gets it done eventually!]
  - Nice extension: What if you prior says $|x_0| \leq 4$ but $(x_0)_{true} = 6$? Explain result.
- Assignment 2 has been released — see Assignment 2.ipynb on Github (in the assignments directory) or from Carmen.
- Mini-project I has been released (but not due for ≈3 weeks). See mini-projects directory. [Quick overview in class.]

Lead-in to Metropolis_Poisson_example.ipynb

- return to Richard McElreath's visualization
- continue with 40 starting with ✱
- Class: work through the notebook
- Things to note (see 43):
  - MCMC trace and warm-up or burn-in
  - Check the scaling with the number of steps ⇒ what improves?

- When you finish (or I tell you!), move on to MCMC-random-walk-and-sampling.ipynb. Discuss the questions with your neighbors.

W(43) 9.

9/13/19

Metropolis Poisson-example from Gregory, section 12.2

- See: mcmc-sampling/Metropolis_Poisson_example.ipynb

- We've already seen the Poisson distribution $p(k|\mu) = \frac{\mu^k e^{-\mu}}{k!}$ $k \geq 0$ integer and we've sampled it through a scipy.stats function. Here we'll do it via MCMC.

- Markov chain: starts with some initial value, then each successive one is generated from previous.

- Step through the procedure for Poisson. (class do this!) etc.
  - Then step through the code.

- Look at the two graphs produced
  - MCMC trace: value at successive MC steps. Notice the fluctuations: it stays reasonably close to 3 but still can jump high.
  - Histogram shows how well we're doing.

  ⇒ use cntr-enter to run many times.
  - Note the outliers at the beginning: needs to equilibrate. This is called the warm-up (or "burn-in") time.
  - How do you expect it to behave for different μ?
  - Do the questions.

- Note: the proposal pdf is asymmetric
  - symmetric means that probability to jump to θ_{new} from θ^t is same as likelihood of jumping back to θ^t from θ_{new}: $q(\theta_{new}|\theta^t)$
    - Typically $N(\theta^t, \sigma)$ with fixed σ.
    - Symmetric because difference of θ_{new}, θ^t appears squared.

9/13/19

Visualization of MCMC Sampling

There are excellent javascript visualizations of MCMC sampling out there.

- A particularly effective set of interactive demos was created by Chi Feng, available at https://chi-feng.github.io/mcmc-demo/
- These demos range from random walk Metropolis Hastings to Adaptive MH to Hamiltonian Monte Carlo to No-U-Turn Sampler (NUTS) to Metropolis-adjusted Langevin Algorithm (MALA) to Hessian-HMC (H2MC), to Stein Variational Gradient Descent (SVGD) to Nested Sampling with RadFriends (RadFriends-NS).

- An accessible introduction to MCMC with simplified versions of Feng's visualization by Richard McElreath. Let's look at the first part of his blog entry at http://elevanth.org/blog/2017/11/28/build-a-better-markov-chain/

- Recall basic structure of Metropolis-Hastings
  1) make a random proposal for new parameter values
  2) accept or reject the proposal based on a Metropolis criterion

- First simulation is Random Walk Metropolis-Hasting
  - Target distribution is two-dimensional Gaussian (just the product)

* If the distribution correlated? How do you know?
  - An arrow indicates a proposal, which is accepted (green) or rejected (red)
  - notice that the direction and a length of the proposal arrow varies.
  - seems to do ok on such a simple distribution, as indicated by how well the projected posteriors get filled in,
  - but it is diffusing – a random walk – which is not so efficient.

  A more complicated shape can cause problems:
  - MH can spend a lot of time exploring over again same regions
  - If not specially tuned, can reject many proposals (red arrows)

9/13/19

- Donut shape is much trickier:
  - Notice that the projected 1d posteriors don't seem to be so complex, but this is a difficult topology.
  - Is it realistic? The claim is that when there are many parameters (high dimensional space), this is analogous to a common target distribution.

- Problems: constantly looking for right step size that is big enough to explore the space, but small enough to not get rejected too much.
  - High dimensions is a big space! Hard to stay in a region of high probability while also exploring enough (in reasonable time).

- Note on donuts in high dimensions
  - see bayes_talk.028.png
  - look at average radius of points sampled from multivariate Gaussians as a function of the dimension
  - blue is 1d, green is 2d, ..., yellow is 6d.
  - imagine yellow as 6 dimensional <u>shell</u> ⇒ <u>analog</u> is two dimensional donut.

- Take a look at Feng site
  - banana distribution – difficult
  - multimodal – very, very tough (see Christian's talks)
  - try adjusting proposal σ (Gaussian proposal with sd = σ)
    - ⇒ try this on donut: to get green you need excellent step size tuning.

- Back to McElreath page. What is the answer? "Better living through physics."
  - This means Hamiltonian Monte Carlo → show examples
  - We'll come back to that in the future and stick with MH for now.