DS4S Group 2 Project Plan
Our slides:
https://docs.google.com/presentation/d/1calZgXMhpjQEFn10a1J_8_c6I7TnlENyz6aRUdxB0Mw/edit?usp=sharing

Problem statement:

1) A reproduction of the Pantheon constraints shown in Fig. 18 of S18, both with and without inclusion of systematic errors.

2) Determine the posterior probability density of H_0 after adding a Gaussian prior on corrected supernova absolute magnitude of M =19.23 +/- 0.042. Make a figure.

Write your own MCMC code (do not use a publicly available code such as CosmoMC).

Grading is based on setup and presentation.

Additional stuff to maybe save for later assignment:

3) Tests for consistency:

A) A residual plot for the binned data similar to Fig. 11 of S18.

B) Evaluation of \chi^2 and calculation of PTE

C) Draw a handful of realizations from the noise covariance, and add to best-fit signal

D) Take the symmetric square root of the error covariance matrix. Transform the residuals r to r' = C^{-1/2}r and histogram. Compare with expectations. Note that <r' r'^T> = C^{-1/2}<r r^T> C^{-1/2} = the identity matrix. So the histogram should be consistent with a Gaussian with unit variance.

E) Break up binned supernova distances into several redshift ranges and test for stability of H_0 inference across these different subsets.

Updated 5/1/20:

Produce the following:

1) A reproduction of the Pantheon constraints shown in Fig. 18 of S18, both with and without inclusion of systematic errors.

2) The posterior probability density of H_0 after adding a Gaussian prior on corrected supernova absolute magnitude of M =-19.23 +/- 0.042 that comes from calibrating corrected supernova peak brightnesses with Cepheids.

Repo Structure:
Modules are each of our first names
(parameters_adam, functionality_pritom, functionality_junying)

Separate directory (test) for tests
Simple “runme”-style file in root directory

Tasks:

Task 1) Proposing parameter set in 3D: Adam

Name: get_new_parameters()
Inputs: sigmas [sigma_h, sigma_m, sigma_v], previous parameters [H, omega_m, omega_v]
(Optional: distribution type {gaussian etc})
Outputs: List of 3 parameter values [H, omega_m, omega_v]

Task 2) Compute likelihood : Junying

Inputs: List of 3 parameter values [H, omega_m, omega_v]
Transients: least-squares errors np.lstsq(\vec{m_b}-\vec{mu})
Outputs: real number ; acceptance_probability

Task 3) Generate MCMC chain : Pritom

Inputs: chain length, number of data points / all the available data, starting state or initial proposals

Supporting functions (dependencies): Proposal generator (from Adam), posterior likelihood calculator (from Junying), data generator (from Junying or I should write the code ?)

Outputs: samples containing the parameter set as an array with dimension (sample number * parameter number)

Subtask : 1. Marginalization or not ?
2. Create data from posterior chain to use for visualization

Task 4) Confirm convergence: Pritom

Inputs: chain
Supporting functions (dependencies) : None; internal test
Outputs: test result

Task 5) Visualization: Adam
show_final_plot()
Visualization (isolate 1 variable, plot 2-d probability density) :

Task 6) Test :

Presentation: towards a **physics colloquium audience**

Your deliverables are the presentation (on May 7th), and the slides you use for the presentation (also due May 7th). The first slide in your slide deck should state the audience of your group, and also state what you want the audience to experience and take away. This first slide is not a part of the presentation to your pretend audience -- it's there so we know what your goal is.

It is up to your group to select what the desired impact (experience and takeaway) is on your audience during these 15 minutes.

Most people over-estimate what an audience can take away and the amount of material that can be covered.

You'll be graded on:

1) worthiness of goal -- did you find something that would capture the audience's interest? (10)

2) appropriateness of amount of information to convey -- most people have trouble keeping this appropriately small (10)

3) clarity of presentation -- did the ideas come across clearly (10)

4) delivery quality -- was their eye contact, speaking voice at an attention-commanding volume, a lack of grasping for the right words (10)

5) audience engagement -- would the audience's attention be maintained throughout? questions to the audience can help. Use of story can help (think of Miller's workshop). Giving the audience the context that makes your work interesting can be very important here too. You need to give the audience reason to pay attention right up front. (20)

Note that it is up to your group to decide which team member(s) are delivering the presentation.

Schedule (Tentative):
Presentation Deadline: May 7th
Presentation Format: Google slides / libra

Task deadlines:
First drafts of all above tasks by Thursday, May 1, 5 PM
Final drafts by Saturday (synchronous meeting to work out kinks?)
Presentation draft by Monday (sections?)
Final presentation draft Tuesday night
Meet Wednesday for trial run
Present Thursday!

Questions (Answered 4/27/20 in office hours):

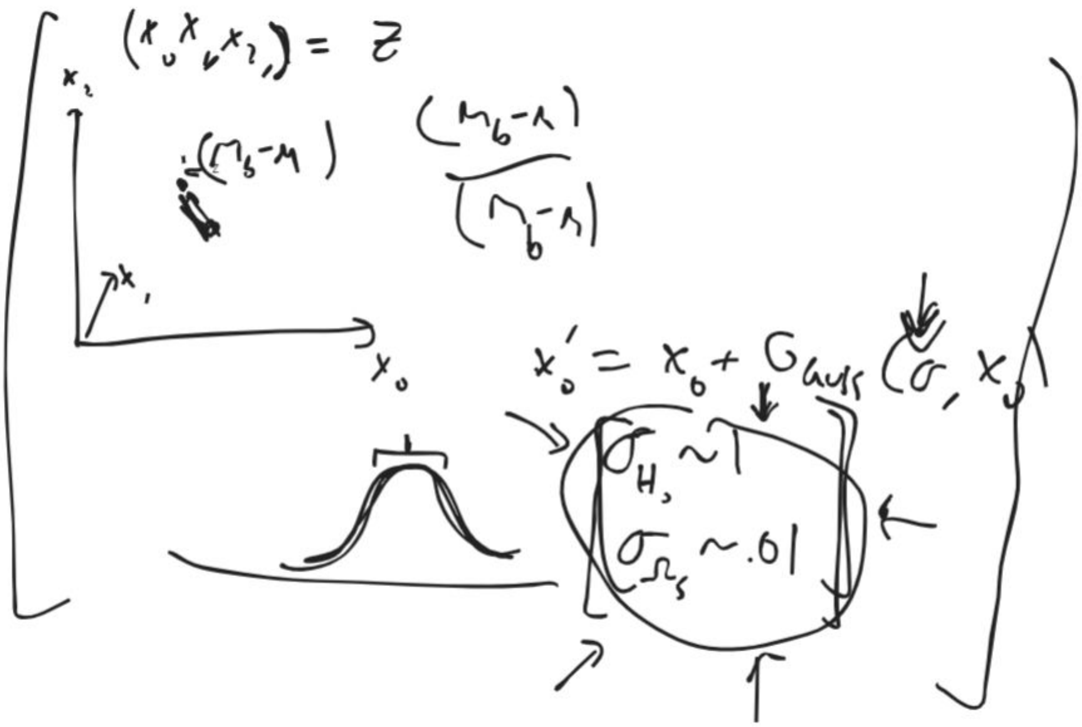
1. Should we use D_l(z) presented in https://iopscience.iop.org/article/10.3847/1538-4357/aab9bb/pdf#%FE%FF%00b%00m%00_%00a%00p%00j%00a%00a%00b%009%00b%00b%00e%00q%00n%003 P 16, or some other form? What is Ωk?
   Prof. Knox: https://en.wikipedia.org/wiki/Angular_diameter_distance
   Also see Adam's handwritten notes: w=-1, but Ωk satisfies Ωk+ Ωm+ Ωlambda = 1
2. What other systematic errors we need to consider besides d_mb ?
   Prof. Knox: Systematic errors! (having to do with smoothing!)
3. Do we care about "Nuisance" parameter M?
   Prof. Knox: Yep, still counts--M is a 4th parameter (can constrain omegas via sum(omegas)=1).
   M and H_0 influence the data the same way (H_0 can be absorbed into a constant sum)
4. Should we impose constraints on omega_m and omega_v ? Or we should consider them as independent ?
   (See above)
5. Should we also consider H0 dependent on omegas' ?
   H_0 should be varied independently.
6. Is our proposed likelihood valid or not ?
   No. Instead, use
   $$P(m_0, m_1, ...m_{40}|\theta, M) \propto \exp\{-[m_i - (\mu(\theta, z_i) - M)](C^{-1})_{ij}[m_j - (\mu(\theta, z_j) - M)]/2\}$$
   Per slides #8.
7. What are the prior distributions for the parameters ?
   Uniform priors for everybody else.
8. Can we adaptively change sigma_h etc, or will this violate some assumption of Markov chain applicability? (Probably no?)
   Should be no need
9. P8 of slides: What's the purpose of the covariance matrix? (something with error?)
   See 2) above and slide 8 of the GroupProjectIntro.key

Useful Resources:
OBSERVATIONAL EVIDENCE FROM SUPERNOVAE FOR AN ACCELERATING UNIVERSE AND A COSMOLOGICAL CONSTANT

The Complete Light-curve Sample of Spectroscopically Confirmed SNe Ia from PanSTARRS1 and Cosmological Constraints from the Combined Pantheon Sample

Random doodle space:

Discussion of navigating 3-D parameter space and relevant sigmas:



To reproduce graphic: align an ellipse with semimajor and major axes pointing to maximize data capture while minimizing least-squares distance to nearest edge? Then scale the axes to get 68% of the data etc.

Presentation options:

Message: Everybody wants to estimate parameters. Here’s one example of how we did it for our application: estimating these cosmo parameters.

Break Data, Model, Parameters on early slide
Our example: These cosmo parameters --Pritom
Why do we want to know them. -- Pritom
Here’s what we did (Bayes inference) -- Junying
Here’s what we did (Markov chain construction) -- Adam
Figures/we did it -- Adam