9/20/19

Physics 8805: Learning from Data: Lecture 9

Notebooks for today:

- mcmc-random-walk-and-sampling.ipynb
- why_bayes_is_better_I.ipynb and bayes_billiard.ipynb
- parameter_estimation_fitting_straight-line_II.ipynb

- On board: Recap points on MCMC
  - want $p(\vec{\theta}|D) = \frac{1}{Z}\underbrace{p(D|\vec{\theta})\,p(\vec{\theta})}_{f(\vec{\theta})}$ ← mcmc samples this!
  - Histogram of samples looks like pdf. Expectation values from samples:

    for parameters: $E_{p(\vec{\theta})}[\vec{\theta}] \approx \frac{1}{N}\sum_{k=1}^{N}\vec{\theta}_k$

    for general $g(\vec{\theta})$ $E_{p(\vec{\theta})}[g(\vec{\theta})] \equiv \dfrac{\int d\vec{\theta}\, g(\vec{\theta}) f(\vec{\theta})}{\int d\vec{\theta}\, f(\vec{\theta})} \approx \frac{1}{N}\sum_{k=1}^{N} g(\vec{\theta}_k)$

  - Autocorrelation

$$\rho(h) = \frac{\sum_{\text{overlap}}\left[(X_t-\bar{X})(X_{t+h}-\bar{X})\right]}{\sqrt{\sum_{\text{overlap}}(X_t-\bar{X})^2}\,\sqrt{\sum_{\text{overlap}}(X_{t+h}-\bar{X})^2}}$$

often $\rho(h) \sim e^{-h/\tau}$ for $h < 2\tau$

X_t is sample at t
X_{t+h} is sample at $t+h$
$h \equiv$ "lag"

Continue with autocorrelation: 49 and mcmc-random-walk... notebook

Basic idea: Given a sequence $Y_t = \{Y_0, Y_1, \ldots\}$ with mean zero, if uncorrelated, $Y_t \cdot Y_{t+h}$ for $h>0$ should be equally positive and negative as we average over h. If mean is not zero, then subtract it first.

$\rho(h)$ calculates $\dfrac{\langle Y_t Y_{t+h}\rangle}{\sqrt{\langle Y_t^2\rangle\langle Y_{t+h}^2\rangle}}$, so $\rho(h)$ is 1 (fully correlated) and if uncorrelated, $\rho(h) \to 0$. In practice fluctuates around zero.

There will be a typical time until $\rho(h)$ fluctuates about $0 \Rightarrow$ autocorrelation time.

- Return in the future to look at autocorrelation and related diagnostics.

9/20/19

Why Bayes is Better I: Nuisance parameters (II)

- Here using parameter_estimation_fitting_straight_line_II.ipynb

- We are back to the problem of fitting a straight line, this time for a real case: velocities (y) and distances (x) for a set of galaxies.
  - A constant standard deviation of $\sigma = 200$ km/sec is given for the y values and no error on x.

- The question: What value and error for the Hubble constant, assuming we know that a straight line is a valid model? We don't care about the intercept; indeed, the model is $v = H_0 x$. ← slope

- Compare three estimates in the notebook:
  i) Maximum likelihood estimate
  ii) Single-parameter inference
  iii) Full Bayesian analysis.

i) log likelihood: $\log p(D|\theta) = -\frac{1}{2}\sum_{i=1}^{N}\left(\log(2\pi\varepsilon_i^2) + \frac{[y_i - y_m(x_i;\theta)]^2}{\varepsilon_i^2}\right)$

- Use scipy optimize, with covariance matrix for errors (more later).
- $\theta = [b, H_0]$

  Result: $b = -26.7 \pm 136.6$ $H_0 = 80.5 \pm 7.4$

ii) Since we don't care about b, maybe should fix it to maximum likelihood estimate (MLE), leaving a one parameter problem.
  - We calculate the likelihood given this value, and then the 68% estimate:

  $H_0 = 80.5 \pm 3.8$ ← underestimates slope uncertainty, because we have assumed intercept is known precisely

- Now we do the Bayesian way!

9/20/19

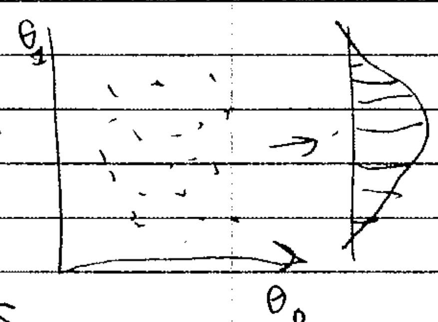
- For priors, we take the symmetric (scale invariant) prior for the slope (recall earlier discussion) and a normal distribution with $\sigma = 200$ for the intercept.
  - Same likelihood.

- Use emcee to do the sampling
  - What happened in the posterior plot? $\Rightarrow$ chains take a while to converge $\Rightarrow$ warm-up or burn-in time.
  - Look at the traces of individual MCMC chains for b and m.
  - Choose warm-up skip of 200 to be conservative.
  - Plot 1, 2, 3 sigma levels.

- Marginalization is simple with MCMC

Want $p(\theta_1 | D, I) = \int d\theta_0 \, p(\theta_0, \theta_1 | D, I)$ given N MCMC samples

output of N samples $\rightarrow$
$$\begin{pmatrix} \theta_{0,1} & \theta_{1,1} \\ \theta_{0,2} & \theta_{1,2} \\ \vdots & \\ \theta_{0,N} & \theta_{1,N} \end{pmatrix} \xrightarrow[\text{2nd column}]{\text{keep}} \begin{pmatrix} \theta_{1,1} \\ \theta_{1,2} \\ \vdots \\ \theta_{1,N} \end{pmatrix} \text{ or }$$



$\Rightarrow$ left side is just $\sum_{i=1}^{N} \theta_{1,i}$!

slope_samples = emcee_trace[1,:]

Result $H_0 = 78.23 + 6.82 - 6.27 \Rightarrow 78.2 + 6.8 - 6.3$ (also corner)

Notebook shows how to make 1σ bands!

Summary:
MLE (1σ): 80.5 +/- 7.4
Fixed intercept (1σ): 80.5 +/- 3.8
Full Bayesian (68% credible region): 78.2 (+6.8, -6.3)

cf. G.B. King et al., PRL 122, 232502 (2019) for Nuclear reactions comparison

9/20/19

Error propagation: Given H_0 with uncertainties or a full posterior as just determined, how do we determine the error in astronomical distances in applying

$$v = H_0 x \,?$$

- H_0 sets the "ruler", so any error propagates. How do we calculate it?
  - More precisely for us Bayesians, given

    $$v_{measured} = (100 \pm 5) \times 10^3 \text{ km/sec}$$

    and the H_0 posterior, what is the posterior pdf for the distance to the galaxy?

- Do this two ways
  1. Fixed value of H_0 from mean of previous analysis
  2. Using the full sampled posterior.

Model: $v_{measured} = v_{theory} + \delta v_{exp}$

where $v_{theory} = H_0 x$ and $\delta v_{exp} \sim N(0, \sigma_v)$, here $\sigma_v = 5 \times 10^3 \frac{km}{s}$

- Assume measurement error in v (i.e. δv_{exp}) is uncorrelated with the error in H_0.

Case 1: fixed H_0

- Prior for x: uniform in $[x_{min}, x_{max}]$

$$\Rightarrow p(x|D,I) \overset{\text{Bayes!}}{\propto} p(D|x,I)\,p(x|I)$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma_v} e^{-\frac{(v_{meas}-v_{th})^2}{2\sigma_v^2}} \; p(x|I)$$

with $H_0 = \hat{H}_0$ fixed.

9/20/19

Case 2: Using the inferred pdf for $\hat{H}_0$

$\Rightarrow$ need to introduce information on H_0.
- This may be a function, or just a set of samples $\{H^{(i)}\}_{i=1}^{N}$ generated by our MCMC sampler.
- As always, we use marginalization (and product rule, Bayes' rule, etc)

$$p(x|D,I) = \int dH_0 \, p(x, H_0 | D, I) \quad \text{and we know } H_0 \text{ independent of } x$$

Bayes rule
$$\propto \int dH_0 \, p(D|x, H_0, I) \, p(x, H_0 | I)$$

independence of x, H_0
$$\propto \int dH_0 \, p(D|x, H_0, I) \, p(x|I) \, p(H_0|I)$$

$$\propto p(x|I) \int dH_0 \, p(H_0|I) \, p(D|x, H_0, I)$$

here assume 1 given the likelihood — sum over marginalized — given in Case 1

from H_0 samples $\Rightarrow$
$$p(x|D,I) \approx \frac{1}{N} \sum_{i=1}^{N} p(D|x, H_0^{(i)}, I)$$
$\leftarrow$ expectation value of $p(D|x, H_0, I)$

$H^{(i)}$ sampled

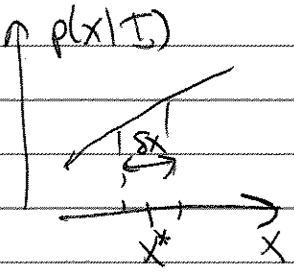
- See notebook for comparison

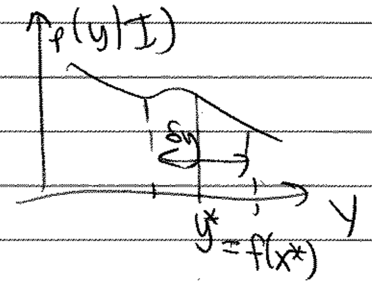
59

9/20/19

Error propagation to functions of uncertain parameters.

That is, if we have a posterior for X, what is the posterior for $Y = f(X)$?





What do we know? Probability in the interval must be the same, regardless of what variable is used, X or Y.

$$\Rightarrow \quad p(X = x^* | I)\,\delta X = p(Y = Y^* | I)\,\delta Y \quad \text{with } Y^* = f(X^*)$$

Must be true for all x^* $\Rightarrow$ $p(X|I) = p(Y|I) \times \left|\frac{dY}{dX}\right|$ in $\delta x, \delta y \to 0$ limit.

Bragg peak example. Signal peak measured as $A = A_0 \pm \sigma_A$

likelihood $\Rightarrow$ $p(D|A,I) \propto e^{-(A-A_0)^2/2\sigma_A^2}$

prior $\Rightarrow$ $p(A|I) = \begin{cases} \text{constant} & 0 \le A \le A_{max} \\ 0 & A < 0 \end{cases}$ physically, amplitude > 0, even if $A_0 < 0$!

The complex structure factor f from $A = |F|^2 \equiv f^2$. What is f?

$$p(f|D,I) = p(A|D,I)\left|\frac{\partial A}{\partial f}\right| \quad {}^{= 2f}$$

$$\Rightarrow p(f|D,I) \propto f\, e^{-(f^2 - A_0)^2/2\sigma_A^2} \quad \text{for } f \geq 0$$

If a Gaussian approximation is good, what is $f_0 \pm \sigma_f$?

$$L(f) = L(f_0) + \frac{1}{2}\underbrace{\frac{d^2L}{df^2}\bigg|_{f_0}}_{= -1/\sigma_f^2}(f-f_0)^2 + \ldots \quad \frac{dL}{df}\bigg|_{f_0} = 0 \Rightarrow \boxed{f_0 = \sqrt{A_0}} \Rightarrow \frac{d^2L}{df^2}\bigg|_{f_0} = \frac{-4A_0}{\sigma_A^2} \Rightarrow \boxed{\sigma_f = \frac{\sigma_A}{2\sqrt{A_0}}}$$

See how this can break down in the why_bayes_is_better_I.ipynb notebook.