

# MATH 390.03-02 / 650 Spring 2016 Homework #10

Professor Adam Kapelner

Due 8:30AM, Wednesday, May 25, 2016

(material covered on exam, writeup not required – for extra credit only)

(this document last updated Monday 16<sup>th</sup> May, 2016 at 10:12pm)

## Instructions and Philosophy

The path to success in this class is to do many problems. Unlike other courses, exclusively doing reading(s) will not help. Coming to lecture is akin to watching workout videos; thinking about and solving problems on your own is the actual “working out.” Feel free to “work out” with others; **I want you to work on this in groups.**

Reading is still *required*. For this homework set, read about Gibbs sampling and Metropolis-Hastings sampling. Also read ch17 and the epilogue in McGrayne.

The problems below are color coded: **green** problems are considered *easy* and marked “[easy]”; **yellow** problems are considered *intermediate* and marked “[harder]”, **red** problems are considered *difficult* and marked “[difficult]” and **purple** problems are extra credit. The *easy* problems are intended to be “giveaways” if you went to class. Do as much as you can of the others; I expect you to at least attempt the *difficult* problems.

Problems marked “[MA]” are for the masters students only (those enrolled in the 650 course). For those in 390, doing these questions will count as extra credit.

This homework is worth 100 points but the point distribution will not be determined until after the due date. See syllabus for the policy on late homework.

Up to 10 points are given as a bonus if the homework is typed using L<sup>A</sup>T<sub>E</sub>X. Links to installing L<sup>A</sup>T<sub>E</sub>X and program for compiling L<sup>A</sup>T<sub>E</sub>X is found on the syllabus. You are encouraged to use [overleaf.com](http://overleaf.com). If you are handing in homework this way, read the comments in the code; there are two lines to comment out and you should replace my name with yours and write your section. The easiest way to use overleaf is to copy the raw text from hwxx.tex and preamble.tex into two new overleaf tex files with the same name. If you are asked to make drawings, you can take a picture of your handwritten drawing and insert them as figures or leave space using the “\vspace” command and draw them in after printing or attach them stapled.

The document is available with spaces for you to write your answers. If not using L<sup>A</sup>T<sub>E</sub>X, print this document and write in your answers. I do not accept homeworks which are *not* on this printout. Keep this first page printed for your records.

NAME: \_\_\_\_\_

## Problem 1

These are questions about McGrayne's book, chapter 17 and the Epilogue. They are optional.

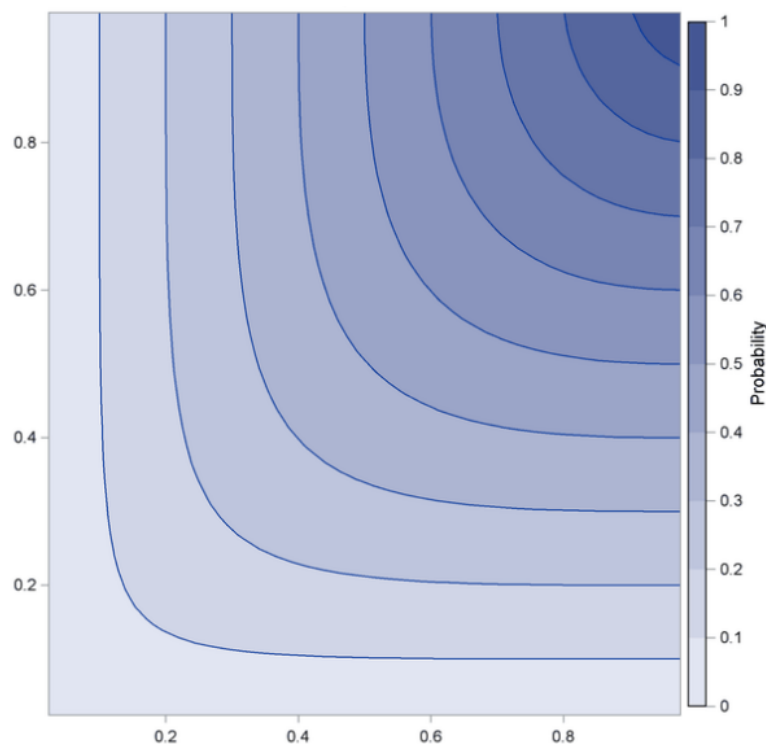
- (a) [easy] What do the computer scientists who adopted Bayesian methods care most about and whose view do they subscribe to? (p233)
- (b) [easy] How was "Stanley" able to cross the Nevada desert?
- (c) [easy] What two factors are leading to the "crumbling of the Tower of Babel?"
- (d) [harder] Does the brain work through iterative Bayesian modeling?
- (e) [easy] According to Geman, what is the most powerful argument for Bayesian Statistics?

## Problem 2

These are questions which introduce Gibbs Sampling.

(a) [easy] Outline the systematic sweep Gibbs Sampler algorithm below (in your notes).

(b) [easy] Pretend you are estimating  $\mathbb{P}(\theta_1, \theta_2 \mid X)$  and the joint posterior looks like the picture below where the  $x$  axis is  $\theta_1$  and the  $y$  axis is  $\theta_2$  and darker colors indicate higher probability. Begin at  $\langle \theta_1, \theta_2 \rangle = \langle 0.5, 0.5 \rangle$  and simulate 5 iterations of the systematic sweep Gibbs sampling algorithm by drawing new points on the plot (just as we did in class).



(c) [harder] We previously have shown that if  $X \mid \theta \sim \text{Binomial}(n, \theta)$  and the prior on  $\theta \sim \text{Beta}(\alpha, \beta)$ , then  $X \sim \text{BetaBinomial}(n, \alpha, \beta)$ . Even though we proved this result, pretend like you didn't know it and create a Gibbs sampler which finds  $\mathbb{P}(X)$ .

(d) [easy] Consider the simple linear regression model under OLS with  $n$  data points. Assume the true  $\beta_0 = 2$  and  $\beta_1 = -2$  and  $\sigma^2 = 1$ . Let's sample  $n = 3$  points. Run the code on lines 1–20 of the code at the link here by copying them from the website and pasting them into an R console. This will plot the 3 data points and provide a best fit line. Do you think the best fit line does a good job?

(e) [easy] Consider the model in (d). We impose priors of  $\beta_0 \sim \mathcal{N}(0, \sigma^2/m)$ ,  $\beta_1 \sim \mathcal{N}(0, \sigma^2/m)$  and  $\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0\sigma_0^2}{2}\right)$  which is the model we did in class. Find the kernel of the posterior.

- (f) [easy] Consider the model in (d). Find the conditional distribution of  $\sigma^2$ . The other two conditionals are filled in below so you don't have to complete the squares:

$$\beta_0 \mid \beta_1, \sigma^2, \mathbf{X}, \mathbf{y} \sim \mathcal{N}\left(\frac{\bar{y}n - \beta_1 \bar{x}n}{n + m}, \frac{\sigma^2}{n + m}\right)$$

$$\beta_1 \mid \beta_0, \sigma^2, \mathbf{X}, \mathbf{y} \sim \mathcal{N}\left(\frac{\sum x_i y_i - \beta_0 \bar{x}n}{m + \sum x_i^2}, \frac{\sigma^2}{m + \sum x_i^2}\right)$$

$$\mathbb{P}(\sigma^2 \mid \beta_0, \beta_1, \mathbf{X}, \mathbf{y}) \propto$$

- (g) [easy] Let's run 10,000 iterations of the Gibbs sampler you created in part (f) by running lines 22–78 of the code at the link here. Explain the three plots that popped up. Where would you say the chains burned in at?

- (h) [easy] Now let's assess autocorrelation for the Gibbs chains in part (g) by running lines 82–87 of the code at the link here. Explain these three plots? What do we mod our chains by to thin them out so the chains represent independent samples?

- (i) [easy] Now run lines 99 – 119 of the code at the link here which will first burn and thin the chains. Explain these three plots. What distributions do these frequency histograms approximate? You must have  $\mathbb{P}(\text{something})$  in your answer. What are the blue lines? What are the red lines? What are the grey lines? Read the code if you have to for the answers.

### Problem 3

These are questions which deal with the change point model. We did this in class in Lecture 23.

- (a) [easy] Consider the change point Poisson model we looked at in class. We have  $m$  exchangeable Poisson r.v.'s with parameter  $\lambda_1$  followed by  $n - m$  exchangeable Poisson r.v.'s with parameter  $\lambda_2$ . Both rate parameters and the value of  $m$  are unknown so the parameter space is 3-dimensional. Write the likelihood below.
- (b) [easy] Consider the model in (a) where  $\lambda_1 = 2$  and  $\lambda_2 = 4$  and  $m = 10$  and  $n = 30$ . Run the code on lines 1–12 of the code at the link here by copying them from the website and pasting them into an R console. This will plot a realization of the data with those parameters. Can you identify the change point visually?
- (c) [easy] Consider the model in (a) but we are blinded to the true values of the parameters given in (b) and we wish to estimate them via a Gibbs sampler. Run the code on lines 19–77 of the code at the link here which will run 10,000 iterations. What iteration number do you think the sampler converged?

- (d) [easy] Now we wish to assess autocorrelation among the chains from the Gibbs sampler run in (d). Run the code on lines 81–86 of the code at the link here. What do we modify our chains by to thin them out so the chains represent independent samples?
- (e) [easy] Run the code on lines 98–118 of the code at the link here which will first burn and thin the chains. Explain these three plots. What distributions do these frequency histograms approximate? You must have  $\mathbb{P}(\text{something})$  in your answer. What are the blue lines? What are the red lines? What are the grey lines? Read the code if you have to for the answers.
- (f) [difficult] Test the following hypothesis:  $H_0 : m \leq 15$  by approximating the  $p$ -value from one of the plots in (e).
- (g) [difficult] [M.A.] Explain a procedure to test  $H_0 : \lambda_1 = \lambda_2$ . You can use the plots if you wish, but you do not have to.

(h) [difficult] What exactly would come from  $\mathbb{P}(X^* | X)$  in the context of this problem? Assume  $X^*$  is the same dimension of  $X$  (in our toy example,  $n = 30$ ). Explain in full detail. Be careful!

(i) [difficult] Explain how you would estimate  $\text{Cov}[\lambda_1, \lambda_2]$  and what do you think this estimate will be close to?

#### Problem 4

These are questions which deal with using the Metropolis-Hastings (M-H) algorithm to estimate the parameters in a Poisson regression. We did this in class in Lecture 23.

(a) [easy] Consider the Poisson regression where we have  $Y_i \overset{exh}{\sim} \text{Poisson}(\beta_0 + \beta_1 t_i)$  with  $n$  samples. Each sample has a timestamp and a count i.e.  $\langle t_i, y_i \rangle$ . Write the likelihood below.

(b) [easy] Explain why you cannot create a systematic sweep Gibbs sampler for  $\mathbb{P}(\beta_0, \beta_1 | X)$ .



- (c) [easy] If you cannot use a Gibbs sampler for  $\mathbb{P}(\beta_0, \beta_1 \mid X)$ , what algorithm can you use instead?
- (d) [easy] Consider the model in (a) where  $\beta_0 = 1$  and  $\beta_1 = 3$  and  $n = 50$ . Run the code on lines 1–22 of the code at the link here by copying the lines (exactly) from the website and pasting them into an R console. This will plot  $n$  realizations of the data with  $t_i$  evenly spaced in  $[0, 2]$ . Can you identify the expectation's linear relationship with time visually?
- (e) [easy] Consider the model in (a) but we are blinded to the true values of the parameters given in (c) and we wish to estimate them via a M-H sampler. Run the code on lines 23–83 of the code at the link here which will run 100,000 iterations. What iteration number do you think the sampler converged?
- (f) [easy] Now we wish to assess autocorrelation among the chains from the Gibbs sampler run in (d). Run the code on lines 87–91 of the code at the link here. What do we modify our chains by to thin them out so the chains represent independent samples?
- (g) [difficult] Why is the autocorrelation so much higher than what you found in the normal-ridge with unknown variance model's Gibbs sampler and the Poisson change point model's Gibbs sampler (problems 2h and 3d)?

- (h) [easy] Run the code on lines 100–115 of the code at the link here which will first burn and thin the chains. Explain these three plots. What distributions do these frequency histograms approximate? You must have  $\mathbb{P}(\text{something})$  in your answer. What are the blue lines? What are the red lines? What are the grey lines? Read the code if you have to for the answers.
- (i) [easy] Did the M-H sampler “capture” the true parameter values in the approximate 95% credible region?
- (j) [easy] Run a hypothesis test of  $H_0 : \beta_1 \leq 2.5$  by approximating the  $p$ -value from the plot.
- (k) [difficult] Explain carefully how you would use MCMC to create samples from  $\mathbb{P}(Y^*(t = 1) \mid Y, t)$ .

- (l) [difficult] Look at lines 41 and 53 in the code here. What is the M-H “candidate density?” You need to reference line 2 as well.
- (m) [easy] Run the code on lines 117–120 of the code at the link here which will print out the acceptance probability for the  $\beta_{t+1,0} \mid \beta_{t,0}$  proposal followed by the acceptance probability of the  $\beta_{t+1,1} \mid \beta_{t,1}$  proposal. Write them down here.
- (n) [difficult] [M.A.] During the times the proposals were not accepted, what exactly was going on?
- (o) [difficult] [M.A.] Why is the candidate density you wrote in (j) silly given the model in (a)? Think carefully about what a Poisson r.v. is and you will see why.