MATH 341 / 650.3 Spring 2020 Homework #8

Professor Adam Kapelner

Due by email, Tuesday 11:59PM, May 19, 2020

(this document last updated Wednesday 13th May, 2020 at 5:14pm)

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 the semi-conjugate normal-inverse-gamma model, grid sampling, Gibbs sampling, and Bayes Factors. Finish the McGrayne book.

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 attempt the *difficult* problems.

Problems marked "[MA]" are for the masters students only (those enrolled in the 650.3 course). For those in 341, 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 LaTeX. Links to instaling LaTeX and program for compiling LaTeX is found on the syllabus. You are encouraged to use `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 LaTeX, 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.

(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?

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	$\texttt{qbeta}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
betabinomial	$\texttt{qbetabinom}(p, n, \alpha, \beta)$	$\texttt{d-}(x, n, \alpha, \beta)$	$\texttt{p-}(x, n, \alpha, \beta)$	$\texttt{r-}(n, \alpha, \beta)$
binomial	$\texttt{qbinom}(p, n, \theta)$	$\texttt{d-}(x, n, \theta)$	$\texttt{p-}(x, n, \theta)$	$\texttt{r-}(n, \theta)$
exponential	$\texttt{qexp}(p, \theta)$	$\texttt{d-}(x, \theta)$	$\texttt{p-}(x, \theta)$	$\texttt{r-}(\theta)$
gamma	$\texttt{qgamma}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
inversegamma	$\texttt{qinvgamma}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
negative-binomial	$\texttt{qnbinom}(p, r, \theta)$	$\texttt{d-}(x, r, \theta)$	$\texttt{p-}(x, r, \theta)$	$\texttt{r-}(r, \theta)$
normal (univariate)	$\texttt{qnorm}(p, \theta, \sigma)$	$\texttt{d-}(x, \theta, \sigma)$	$\texttt{p-}(x, \theta, \sigma)$	$\texttt{r-}(\theta, \sigma)$
poisson	$\texttt{qpois}(p, \theta)$	$\texttt{d-}(x, \theta)$	$\texttt{p-}(x, \theta)$	$\texttt{r-}(\theta)$
T (standard)	$\texttt{qt}(p, \nu)$	$\texttt{d-}(x, \nu)$	$\texttt{p-}(x, \nu)$	$\texttt{r-}(\nu)$
T (nonstandard)	$\texttt{qt.scaled}(p, \nu, \mu, \sigma)$	$\texttt{d-}(x, \nu, \mu, \sigma)$	$\texttt{p-}(x, \nu, \mu, \sigma)$	$\texttt{r-}(\nu, \mu, \sigma)$
uniform	$\texttt{qunif}(p, a, b)$	$\texttt{d-}(x, a, b)$	$\texttt{p-}(x, a, b)$	$\texttt{r-}(a, b)$

Table 1: Functions from R (in alphabetical order) that can be used on this assignment and exams. The hyphen in colums 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 2

This problem is about the normal-normal model using a "semi-conjugate" prior. Assume $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ throughout.

(a) [easy] If θ and σ^2 are assumed to be independent, how can $\mathbb{P}(\theta, \sigma^2)$ be factored?

(b) [easy] If $\mathbb{P}(\theta) = \mathcal{N}(\mu_0, \tau^2)$ and $\mathbb{P}(\sigma^2) \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$, find the kernel of the joint posterior, $\mathbb{P}(\theta, \sigma^2 \mid X)$.

(c) [difficult] Show that this kernel can be factored into the kernel of a normal where the leftover is *not* the kernel of an inverse gamma. This is in the lecture notes.

(d) [difficult] [MA] Find the posterior mode of σ^2 using $k(\sigma^2 \mid X)$.

(e) [difficult] Describe how you would sample from $k(\sigma^2 \mid X)$. Make all steps explicit and use the notation from Table 1.

(f) [difficult] Describe how you would sample from $\mathbb{P}(\theta, \sigma^2 \mid X)$. Make use of the sampling algorithm in the previous question. Make all steps explicit and use the notation from Table 1.

(g) [difficult] What are the two main disadvantages of grid sampling?

(h) [difficult] Why do you think the prior $\mathbb{P}(\theta) = \mathcal{N}(\mu_0, \tau^2)$ and $\mathbb{P}(\sigma^2) \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$ is called "semi-conjugate"?

(i) [E.C.] [MA] Find the MMSE of σ^2

Problem 3

These are questions which introduce Gibbs Sampling.

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

(b) [E.C.] Under what conditions does this algorithm converge?

(c) [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 θ_1 and the y axis is θ_2 and darker colors indicate higher probability. Begin at $[\theta_1, \theta_2] = [0.5, 0.5]$ and simulate 5 iterations of the systematic sweep Gibbs sampling algorithm by drawing new points on the plot (just as we did in class).
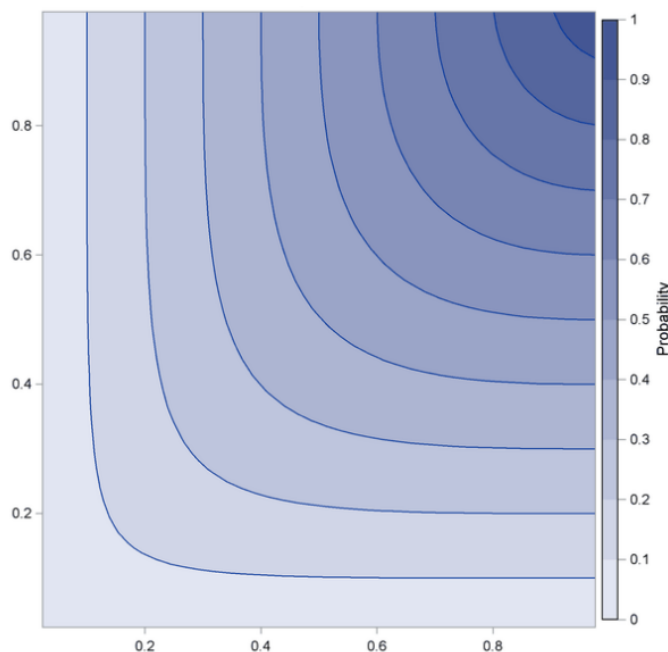
Problem 4

These are questions about the change point model and the Gibbs sampler to draw inference for its parameters. You will have to use R to do this question. If you do not have it installed on your computer, you can use R online without installing anything by using a site like jupyter. You copy code into the black box and click the "run" button atop. Then you enter more code into the next box and click "run" again, etc.

(a) [easy] Consider the change point Poisson model we looked at in class. We have m exchangeable Poisson r.v.'s with parameter λ_1 followed by $n-m$ exchangeable Poisson r.v.'s with parameter λ_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 1–14 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 16–78 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 79–89 of the code at the link here. What do we mod our chains by to thin them out so the chains represent independent samples?

(e) [easy] Run the code on lines 91–121 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 $\mathbb{P}(X^* \mid X)$ 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) [E.C.] Explain how you would estimate $\mathbb{C}\text{ov}[\lambda_1, \lambda_2]$ and what do you think this estimate will be close to?

These are questions about the mixture-of-two-normals model and the Gibbs sampler to draw inference for its parameters. You will have to use R to do this question. If you do not have it installed on your computer you can use this website which will give you provide you with a workable R console.

(a) [easy] Consider the mixture-of-two-normals model we looked at in class. Write the likelihood below.

(b) [easy] Consider the model in (a) with $\theta_1 = 0$, $\theta_2 = 4$, $\sigma_1^2 = 2$, $\sigma_2^2 = 1$ and $\rho = 2$. These are the targets of inference so pretend you don't know their values! Run the code on lines 1–16 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 that it's a mixture of two normals 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 19–92 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 96–103 of the code at the link here. What do we mod our chains by to thin them out so the chains represent independent samples?

9

(e) [easy] Run the code on lines 120–152 of the code at the link here which will first burn and thin the chains. Explain these five plots. What distributions do these frequency histograms approximate?

(f) [difficult] Provide and approximate $CR_{\rho, 95\%}$. Does it capture the true value of ρ?

(g) [difficult] Explain carefully how you would approximate $\mathbb{P}\left(X^* \mid X\right)$.

(h) [difficult] Explain carefully how you would approximate the probability that the 17th observation belonged to the $\mathcal{N}\left(\theta_1, \sigma_1^2\right)$ distribution.

(i) [easy] If one of the θ's did not have a known conditional distribution, which algorithm could you use? Would this algorithm take longer or shorter to converge than the Gibbs sampler you've seen here?

(j) [E.C.] Explain carefully how you would test if $\theta_1 \neq \theta_2$.

Problem 6

This question is about a famous extra sensory perception (ESP) experiment.

According to quantum mechanics, an event should happen with exactly probability 50%. So if someone comes with a claim that they have ESP and can affect this event with their mind, then the event no longer has probability 50% (this is what we are trying to prove).

An experiment was run with someone claiming they have ESP. They tried to affect the event with their mind. The data is as follows: of $n = 104,490,000$ observations, the number of events was $x = 52,263,970$. We will now test to see if the person has ESP a number of ways and try to reconcile the differences.

(a) [easy] What is the MLE for the probability of the event?

(b) [easy] Run a two-sided frequentist test at $\alpha = 5\%$ and report the decision (i.e. retain or reject) and the p value.

(c) [harder] Run a two-sided Bayesian test assuming the principle of indifference. Use the equivalence region approach with $\delta = 0.01$. Report the decision (i.e. retain or reject) and the p value.

(d) [difficult] Calculate the Bayes Factor B where the alternative hypothesis of $\theta \neq 0.5$ can be summed up with $\theta \sim U(0, 1)$.

(e) [difficult] Try to reconcile parts (b), (c), (d), (e). Why are there different answers? What is going on??