MATH 341 / 650.3 Spring 2020 Homework #7

Professor Adam Kapelner

Due by email, Friday 11:59PM, May 8, 2020

(this document last updated Thursday 30th April, 2020 at 8:41am)

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 normal-inverse -gamma / normal-inverse-gamma conjugate model and its marginal posterior distributions and its posterior predictive distribution. Review the general Students T_ν distribution. Read chapter 16 in the McGrayne book - a very long and important chapter.

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 16.

(a) [easy] What was the main problem facing Bayesian Statistics in the early 1980's?

(b) [harder] What is the "curse of dimensionality?"

(c) [easy] How did Bayesian Statistics help sociologists?

(d) [easy] How did Gibbs sampling come to be?

(e) [easy] Were the Geman brothers the first to discover the Gibbs sampler?

(f) [easy] Who officially discovered the expectation-maximization (EM) algorithm? And who *really* discovered it?

(g) [harder] How did Bayesians "break" the curse of dimensionality?

(h) [harder] Consider the integrals we use in class to find expectations or to approximate PDF's / PMF's — how can they be replaced?

(i) [easy] What did physicists call "Markov Chain Monte Carlo" (MCMC)? (p222)

(j) [easy] Why is sampling called "Monte Carlo" and who named it that?

(k) [easy] The Metropolis-Hastings (MH) Algorithm is world famous and used in myriad applications. Why didn't Hastings get any credit?

(l) [easy] The combination of Bayesian Statistics + MCMC has been called ... (p224)

(m) [E.C.] p225 talks about Thomas Kuhn's ideas of "paradigm shifts." What is a "paradigm shift" and does Bayesian Statistics + MCMC qualify?

(n) [easy] How did the BUGS software change the world?

(o) [easy] Lindley said that Bayesian Statistics would win out over Frequentist Statistics because it was more logical. What in reality was the reason for the eventual victory of Bayes?

(p) [E.C.] One of my PhD advisors, Ed George at Wharton told me that "Bayesian Statistics is really 'knowledge engineering."' Is this true? Explain.

(q) [E.C.] Take a look at the software Stan. What kind of potential does it have to change the world? Note: I had an opportunity to work on Stan as a postdoc (right after I finished his PhD) but chose to come to QC instead.

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

Now we will move to the Bayesian normal-normal model for estimating both the mean and variance and demonstrate similarities with the classical results.

(a) [harder] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\ \sigma^2)$ and X represents all $X_1, \ldots, X_n$, Find the kernel of $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ if $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$. Use the substitution that we made in class:

$$\sum_{i=1}^{n}(x_i - \theta)^2 = (n-1)s^2 + n(\bar{x} - \theta)^2$$

where $s^2 := \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$. We do this here because this substitution is important for what comes next.

(b) [harder] Using Bayes Rule, break up $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ into two pieces. How are those two pieces distributed?

(c) [harder] Using your answer from (b), explain in English how you can create samples from the distribution $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ that look like $\{[\theta_1, \sigma_1^2], [\theta_2, \sigma_2^2], \ldots, [\theta_S, \sigma_S^2]\}$.

(d) [difficult] Using these samples, how would you estimate $\mathbb{E}[\theta \mid X]$ and $\mathbb{E}[\sigma^2 \mid X]$? Why is $\mathbb{E}[\theta \mid X]$ of paramount importance?

(e) [difficult] Using these samples, how would you estimate a 95% CR for θ?

(f) [difficult] Using these samples, how would you obtain a p-val for testing if $\sigma^2 > 1.364$?

(g) [difficult] [MA] Using these samples, how would you estimate $\text{Corr}\left[\theta \mid X,\ \sigma^2 \mid X\right]$ i.e. the correlation between the posterior distributions of the two parameters?

(h) [easy] Find $\mathbb{P}\left(\theta \mid X,\ \sigma^2\right)$ by using the full posterior kernel from (a) and then conditioning on σ^2. You should get the same answer as we did before the midterm.

(i) [easy] Find $\mathbb{P}\left(\sigma^2 \mid X,\ \theta\right)$ by using the full posterior kernel from (a) and then conditioning on θ. You should get the same answer as we did before the midterm.

(j) [difficult] Show that $\mathbb{P}\left(\theta \mid X\right)$ is a non-standard T distribution and find its parameters. Assume the prior $\mathbb{P}\left(\theta,\ \sigma^2\right) \propto \frac{1}{\sigma^2}$. The answer is in the notes, but try to do it yourself.

(k) [difficult] Show that $\mathbb{P}(\sigma^2 \mid X)$ is an inverse gamma and find its parameters. Assume the prior $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$. The answer is in the notes, but try to do it yourself.

(l) [easy] Write down the distribution of $\mathbb{P}(X^* \mid X)$ assuming the prior $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$. This is in the notes.

(m) [E.C.] Prove what you wrote in the previous question: $\mathbb{P}(X^* \mid X)$ is the non-standard T distribution and find its parameters.

(n) [harder] Explain how to sample from the distribution of $\mathbb{P}(X^* \mid X)$. Also in the notes.

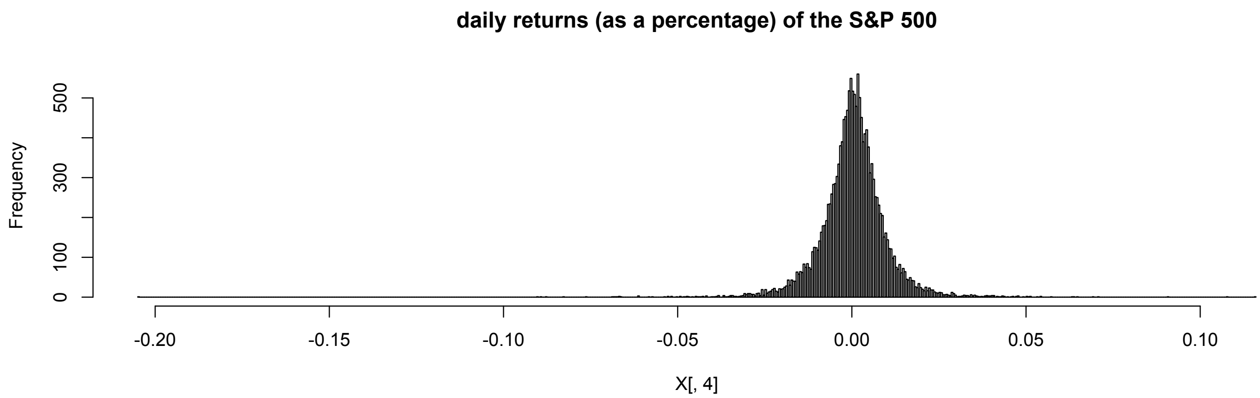
(o) [harder] Now consider the informative conjugate prior of $\mathbb{P}(\theta,\ \sigma^2) = \mathbb{P}(\theta \mid \sigma^2)\,\mathbb{P}(\sigma^2)$ where $\mathbb{P}(\theta \mid \sigma^2) = \mathcal{N}\left(\mu_0,\ \frac{\sigma^2}{m}\right)$ and $\mathbb{P}(\sigma^2) = \text{InvGamma}\left(\frac{n_0}{2},\ \frac{n_0\sigma_0^2}{2}\right)$ i.e. the general normal-inverse-gamma. What is its kernel? Collect common terms and be neat.

(p) [difficult] [MA] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta,\ \sigma^2)$ and given the general prior above, find the posterior and demonstrate it that the normal-inverse gamma is conjugate for the normal likelihood with both mean and variance unknown. This is what I did *not* do in class.

Problem 3

We model the returns of S&P 500 here.

(a) [easy] Below are the 16,428 daily returns (as a percentage) of the S&P 500 dating back to January 4, 1950 and the code used to generate it. Does the data look normal? Yes/no



(b) [harder] Do you think the data is $\stackrel{iid}{\sim}$? Explain.

(c) [harder] Assume $\stackrel{iid}{\sim}$ normal data regardless of what you wrote in (a) and (b). The sample average is $\bar{x} = 0.0003415$ and the sample standard deviation is $s = 0.0096$. Under an objective prior, give a 95% credible region for the true mean daily return.

(d) [difficult] Give a 95% credible region for *tomorrow's* return using functions in Table 1.