MATH 341 / 650.3 Spring 2018 Homework #5

Professor Adam Kapelner

Due Friday 11:59PM, May 4, 2018 under the door of KY604

(this document last updated Monday 23rd April, 2018 at 8:53am)

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, review the Normal-Inverse Gamma, the two-dimensional NormalInverseGamma, the marginal inverse gamma, marginal Students T models, ideas about sampling and read the relevant sections of 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 650.3 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, chapters 15 and 16.

(a) [easy] During the H-Bomb search in Spain and its coastal regions, RAdm. William Guest was busy sending ships here, there and everywhere even if the ships couldn't see the bottom of the ocean. How did Richardson use those useless searches?

(b) [harder] When the Navy was looking for the *Scorpion* submarine, they used Monte Carlo methods (which we will see in class soon). How does the description of these methods by Richardson (p199) remind you of the "sampling" techniques to approximate integrals we did in class?

(c) [harder] What is a Kalman filter? Read about it online and write a few descriptive sentences.

(d) [harder] Where do frequentist methods practically break down? (end of chapter 15)

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

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

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

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

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

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

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

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

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

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

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

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

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

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

(s) [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?

(t) [E.C.] One of my PhD advisors, Ed George at Wharton told me that "Bayesian Statistics is really 'knowledge engineering."' Is this true? Explain.

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	`qbeta`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
betabinomial	`qbetabinom`(p, n, α, β)	`d-`(x, n, α, β)	`p-`(x, n, α, β)	`r-`(n, α, β)
betanegativebinomial	`qbeta_nbinom`(p, r, α, β)	`d-`(x, r, α, β)	`p-`(x, r, α, β)	`r-`(r, α, β)
binomial	`qbinom`(p, n, θ)	`d-`(x, n, θ)	`p-`(x, n, θ)	`r-`(n, θ)
exponential	`qexp`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
gamma	`qgamma`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
geometric	`qgeom`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
inversegamma	`qinvgamma`(p, α, β)	`d-`(x, α, β)	`p-`(x, α, β)	`r-`(α, β)
negative-binomial	`qnbinom`(p, r, θ)	`d-`(x, r, θ)	`p-`(x, r, θ)	`r-`(r, θ)
normal (univariate)	`qnorm`(p, θ, σ)	`d-`(x, θ, σ)	`p-`(x, θ, σ)	`r-`(θ, σ)
poisson	`qpois`(p, θ)	`d-`(x, θ)	`p-`(x, θ)	`r-`(θ)
T (standard)	`qt`(p, ν)	`d-`(x, ν)	`p-`(x, ν)	`r-`(ν)
T (nonstandard)	`qt.scaled`(p, ν, μ, σ)	`d-`(x, ν, μ, σ)	`p-`(x, ν, μ, σ)	`r-`(ν, μ, σ)
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

We now continue questions on the normal-normal conjugate model.

(a) [easy] 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 $\sigma^2 \mid X,\ \theta$. Show that this is the kernel of an inverse gamma. Use the $\hat{\sigma}^2$ substituion we did in class.

(b) [harder] Why is using $\hat{\sigma}^2$ permitted in the setup in (a) but doesn't make sense in the ususal frequentist setup when the likelihood is normal? Hint: what is your target of estimation usually?

(c) [easy] In class we looked at $\sigma^2 \sim \text{InvGamma}(\alpha,\ \beta)$ but we used a different parameterization. Write the different parameterization below and explain why this was done i.e. interpret the meaning of the two new parameters.

(d) [harder] Show that the posterior $\sigma^2 \mid X,\ \theta$ is distributed as an inverse gamma with the prior from (d) and find its parameters.

(e) [easy] What is the Jeffrey's prior for σ^2 (look in the notes and write it down — no need to prove it). Is it proper?

(f) [easy] Show that the Jeffrey's prior for σ^2 is an improper inverse gamma distribution and find its parameters. Note these parameters are not in the parameter space of a proper inverse gamma distribution.

(g) [easy] Under the Jeffrey's prior for σ^2, what is the posterior?

(h) [harder] You are in a milk manufacturing plant producing 1 quart cartons of whole milk. You are willing to assume that the nozzle emits 1 qt on average. In your previous job, you remember inspecting 3 cartons of which you saw 1.02, 0.97, 1.03 quarts of milk inside. Create a prior based on what you've seen in your previous job. This forces you to understand (d).

(i) [difficult] The company wishes to test if there's too much variability i.e. that there is more than $\sigma = 0.1$ variability. You take a sample of 10 and see 1.153, 1.045, 1.268, 1.333, 0.799, 1.075, 1.27, 1.07, 1.192 and 1.079 quarts. Find the p value. You can write the answer below as a function of `rinvgamma`, `qinvgamma` or `pinvgamma` (i.e., expressions from Table 1). E.C. for computing it and testing this at $\alpha = 5\%$. You may want to use the `actuar` package (see here).

(j) [harder] Find $CR_{\sigma^2, 90\%}$ for the data above using expressions from Table 1.

Problem 3

We will review classical *frequentist* concepts from "Math 241/242". Much of this can be drawn from lecture 14 first page.

(a) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ and $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$, what is the distribution of the following:

$$\frac{\bar{X} - \theta}{\frac{\sigma}{\sqrt{n}}} \sim$$

(b) [easy] If $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ and $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$, what is the distribution of $\bar{X}$ assuming σ^2 is known? This can be derived from (a) or found in your Math 241 notes.

(c) [easy] Write the definition of S^2, the r.v. which is the sample variance *estimator*. Hint: use capital letters.

(d) [easy] Write the definition of S, the sample standard deviation *estimator* (or standard error estimator — both terms are synonymous). Hint: use capital letters.

(e) [easy] Write the definition of s^2, the r.v. which is the sample variance *estimate*. Hint: use lowercase letters.

(f) [easy] This answer is in the notes. If $X_1, \ldots, X_n \overset{iid}{\sim} \mathcal{N}(\theta,\, \sigma^2)$ and $\bar{X} := \frac{1}{n}\sum_{i=1}^n X_i$, what is the distribution of the following where S is defined as in (d):

$$\frac{\bar{X} - \theta}{\frac{S}{\sqrt{n}}} \sim$$

(g) [easy] Write the PDF of the general (also called non-standard) T distribution below. You need to use the notation given in class.

(h) [easy] What is the kernel of the nonstandard T?

(i) [harder] What is the distribution of $\bar{X}$ assuming σ^2 is *unknown*? This will differ from (b). Use the answer from part (k) above and the fact that $aT_\nu + c \sim T_\nu(c, a)$ which means that if you shift and scale a T with ν degrees of freedom, you get a nonstandard T_ν with the new center and scaling as parameters.

Problem 4

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) [easy] 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 $\theta,\ \sigma^2 \mid X = x$. 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 is your answer from 2(e). We do this here because this substitution is important for what comes next.

(b) [harder] If $\mathbb{P}(\theta,\ \sigma^2) \propto \frac{1}{\sigma^2}$, show that this is a conjugate prior for the posterior of both the mean and variance, $\mathbb{P}(\theta,\ \sigma^2 \mid X)$. We called this two-dimensional distribution the "normal-inverse-gamma" distribution but we did not go into details about it.

(c) [harder] Using Bayes Rule, break up $\mathbb{P}(\theta,\ \sigma^2 \mid X)$ into two pieces.

(d) [harder] Using your answer from (c), explain how you can create samples $[\theta_s, \sigma_s^2]$ from the distribution $\mathbb{P}(\theta,\ \sigma^2 \mid X)$.

(e) [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?

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

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

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

(i) [harder] Show that $\mathbb{P}(\sigma^2 \mid X)$ is an inverse gamma distribution and find its parameters.

(j) [difficult] Show that $\mathbb{P}(\theta \mid X)$ is a non-standard T distribution (assume prior in b). The answer is in the notes, but try to do it yourself.

(k) [easy] Write down the distribution of $X^* \mid X$ which is in the notes (lec 14, page 6). Note that the answer I wrote down is for the non-informative prior only.

(l) [E.C.] [MA] Prove (k).

(m) [easy] Explain how to sample from the distribution of $X^* \mid X$. Hint: write it as a double integrel of two conditional distributions and a marginal distribution (all conditional on X).

(n) [easy] Now consider the informative conjugate prior of

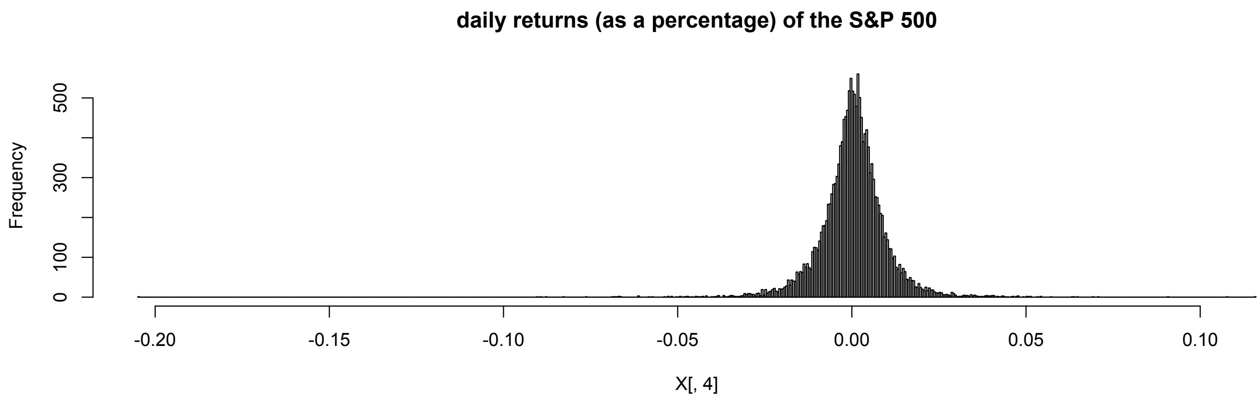
$$\mathbb{P}\left(\theta,\ \sigma^2\right) = \mathbb{P}\left(\theta \mid \sigma^2\right)\mathbb{P}\left(\sigma^2\right) = \mathcal{N}\left(\mu_0,\ \frac{\sigma^2}{m}\right)\text{InvGamma}\left(\frac{n_0}{2},\ \frac{n_0\sigma_0^2}{2}\right).$$

i.e. the general normal-inverse-gamma. What is its kernel? Collect common terms and be neat.

Problem 5

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 $\overset{iid}{\sim}$? Explain.

(c) [harder] Assume $\overset{iid}{\sim}$ 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.