MATH 341 / 650 Spring 2017 Homework #7

Professor Adam Kapelner

Due 2PM under my office door (KY604), Monday, May 8, 2017

(this document last updated Saturday 29^{th} April, 2017 at 10:17pm)

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-Normal model for mean estimation given the variance and variance estimation given the mean and joint mean-variance estimation under the conjugate and semi-conjugate priors. Also, review the concepts of posterior predictive distributions, conjugacy and kernels.

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 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 11–14.

(a) [easy] Did Savage like Shlaifer? Yes / No and why?

(b) [easy] How did Neyman-Pearson approach statistical decision theory? What is the weakness to this approach? (p145)

(c) [easy] Who popularized "probability trees" (and "tree flipping") similar to exercises we did in Math 241?

(d) [easy] Where are Bayesian methods taught more widely than any other discipline in academia?

(e) [easy] Despite the popularity of his Bayesian textbook on business decision theory, why didn't Schlaifer's Bayesianism catch on in the real world of business executives making decisions?

(f) [easy] Why did the pollsters fail (big time) to predict Harry Truman's victory in the 1948 presidential election?

(g) [easy] When does the diference between Bayesianism and Frequentism grow "immense"?

(h) [easy] How did Mosteller demonstrate that Madison wrote the 12 Federalist papers of unknown authorship?

(i) [easy] Write a one paragraph biography of John Tukey.

(j) [easy] Why did Alfred Kinsey's wife want to poison John Tukey?

(k) [easy] Tukey helped NBC with polling predictions for the presidential campaign. What was NBC's polling algorithm based on?

(l) [easy] Why is "objectivity an heirloom ... and ... a fallacy?"

(m) [easy] Why do you think Tukey called Bayes Rule by the name "borrowing strength?"

(n) [easy] Why is it that we don't know a lot of Bayes Rule's modern history?

(o) [easy] Generally speaking, how does Nate Silver predict elections?

(p) [easy] How many Bayesians of import were there in 1979?

(q) [easy] What advice did Chernoff give to Susan Holmes? (Note: Susan Holmes was my undergraduate advisor).

(r) [easy] How did Rasmussen's team estimate the probability of a nuclear plant core meltdown?

(s) [easy] How did the Three Mile Island accident vindicate Rasmussen's committee report?

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-`$(n,\ \alpha,\ \beta)$
betanegativebinomial	`qbeta_nbinom`$(p,\ r,\ \alpha,\ \beta)$	`d-`$(x,\ r,\ \alpha,\ \beta)$	`p-`$(x,\ r,\ \alpha,\ \beta)$	`r-`$(r,\ \alpha,\ \beta)$
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-`$(\alpha,\ \beta)$
geometric	`qgeom`$(p,\ \theta)$	`d-`$(x,\ \theta)$	`p-`$(x,\ \theta)$	`r-`(θ)
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-`$(\theta,\ \sigma)$
normal (multivariate)		`dmvnorm`$(\boldsymbol{x},\ \boldsymbol{\mu},\ \boldsymbol{\Sigma})$		`r-`$(\boldsymbol{\mu},\ \boldsymbol{\Sigma})$
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

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 3

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 $X_1, \ldots, X_n$, in HW6 6(c) you found the kernel of $\theta,\ \sigma^2 \mid X = x$. Do so again below but this time 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 before the midterm.

(i) [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.

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

(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.

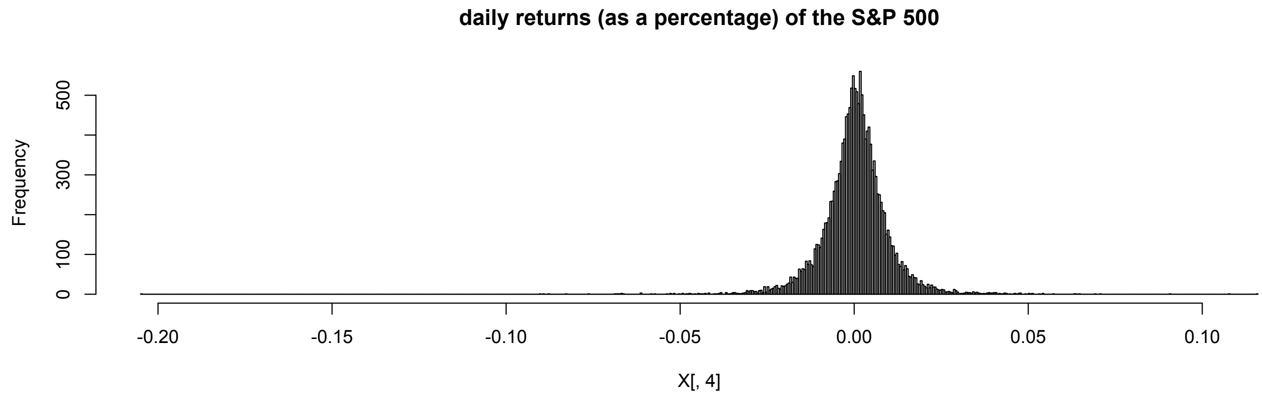
(o) [difficult] [MA] If $X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ and given the general prior above, find the posterior and demonstrate it that the normal-inverse gamma is conjugate for the normal likelihood with both mean and variance unknown. This is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

Problem 4

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

```
X = read.csv('sp_tot_ret_price_1950.csv')
n = nrow(X)
n
hist(X[,4], br = 1000,
  main = 'daily returns (as a percentage) of the S&P 500')
```



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

(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 a relatively objective prior, give a 95% credible region for the true mean daily return.

(d) [difficult] Give a 95% credible region for *tomorrow's* return.

Problem 5

This problem is about the normal-normal model using a "semi-conjugate" prior. Assume $X_1, \ldots, X_n \mid \theta, \sigma^2 \stackrel{iid}{\sim} \mathcal{N}(\theta,\, \sigma^2)$ 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 $\theta \sim \mathcal{N}(\mu_0,\, \tau^2)$ and $\sigma^2 \sim \text{InvGamma}\left(\frac{n_0}{2},\, \frac{n_0\sigma_0^2}{2}\right)$, find the kernel of $\mathbb{P}(\theta,\ \sigma^2)$.

(c) [easy] Using your answer to (b), find the kernel of $\mathbb{P}(\theta,\ \sigma^2 \mid X)$.

(d) [difficult] Show that the kernel in (c) cannot be factored into the kernel of a normal and the kernel of an inverse gamma. This is in the lecture.

(e) [difficult] Your answer to (d) looks like a normal and a $k(\sigma^2 \mid X)$. Find the posterior mode of θ.

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

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

(h) [difficult] What's the bad part about not using a conjugate model?

(i) [E.C.] [MA] Find the MMSE of σ^2.