MATH 390 / 650.3 Spring 2022 Homework #5

Professor Adam Kapelner

Due by email 11:59PM April 28, 2022

(this document last updated Friday 8th April, 2022 at 4:57pm)

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-normal conjugate model and read chapter 15 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 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 15.

(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)

We now discuss the theory of the normal-normal conjugate model. Assume

$$\mathcal{F}: X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\theta, \sigma^2\right) \quad \text{and } \sigma^2 \text{ known}$$

and "X", is the usual shorthand for all $X_1, \ldots, X_n$ samples.

(a) [easy] Show that the kernel of the normal distribution is, $\mathbb{P}\left(X_1 \mid \theta, \sigma^2\right) \propto k(X_1 \mid \theta, \sigma^2) = e^{ax}e^{-bx^2}$ and solve for the values of a and b as functions of θ and σ^2.

(b) [difficult] Assume $\mathbb{P}(\theta) = \mathcal{N}(\mu_0, \tau^2)$. Using the result from (a), show that posterior distribution is normal. Try to do it yourself and only copy from the notes if you must.

(c) [easy] Find the Bayesian point estimates as function of the data and prior's hyperparameters (i.e. $\hat{\theta}_{\text{MMSE}}$, $\hat{\theta}_{\text{MMAE}}$ and $\hat{\theta}_{\text{MAP}}$).

(d) [harder] On a previous homework we showed that if $\mathcal{F} : X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}(\theta, \sigma^2)$ then $\hat{\theta}_{\mathrm{MLE}} = \bar{x}$. Assuming this MLE, show that $\hat{\theta}_{\mathrm{MMSE}}$ is a shrinkage estimator and find ρ.

(e) [harder] Rederive the posterior distribution under Laplace's prior of indifference. This is easier than the exercise in (b) as there are less terms.

(f) [easy] Assuming $\sigma^2 = 1.3$, Laplace's prior and a dataset of $n = 10$ with values 0.48 0.39 1.29 1.02 1.55 -0.22 0.01 -0.52 -1.50 0.71, provide a Bayesian point estimate.

(g) [easy] Assuming the prior, σ^2 and the dataset from (f), provide a 95% CR for θ.

(h) [easy] Assuming the prior, σ^2 and the dataset from (f), provide notation that calculates the p value for a test of $\theta > 1$.

(i) [harder] Rederive Jeffrey's prior. Is it proper?

(j) [harder] Now let $\tau^2 := \sigma^2/n_0$ which is a reparameterization from τ^2 to n_0. Substitute this change into the posterior distribution from (b) to derive the posterior distribution under this reparemterization.

(k) [easy] Provide pseudocount interpretations of μ_0 and n_0.

(l) [easy] Using the pseudocount interpretations of μ_0 and n_0, what is Haldane's prior of ignorance? Is it proper?

(m) [easy] If the prior is $\mathbb{P}(\theta) = \mathcal{N}(\mu_0, \sigma^2/n_0)$, write the integral that will compute the posterior predictive distribution $\mathbb{P}(X_* \mid X, \sigma^2)$ when $n_* = 1$.

(n) [difficult] Compute the posterior predictive distribution $\mathbb{P}(X_* \mid X, \sigma^2)$ when $n_* = 1$ using the integral from (m). Try to do it yourself and if you get stuck, look in the notes. Note the change of parameterization from the notes!

Problem 3

We now discuss the theory of the normal-inverse-gamma conjugate model. Assume

$$\mathcal{F} : X_1, \ldots, X_n \mid \theta, \sigma^2 \overset{iid}{\sim} \mathcal{N}\left(\theta, \sigma^2\right) \quad \text{and } \theta \text{ known}$$

and "X", is the usual shorthand for all $X_1, \ldots, X_n$ samples.

(a) [easy] Find the kernel of $\mathbb{P}(X \mid \theta, \sigma^2)$.

(b) [easy] Assume $\mathbb{P}(\sigma^2) \propto 1$ and then find the kernel of $\mathbb{P}(\sigma^2 \mid X, \theta)$. Substitute in the MLE for σ^2.

(c) [easy] Time for proving the calculus exercise that I always skip. Prove that

$$\int_0^\infty y^{\alpha-1} e^{-\beta y} = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

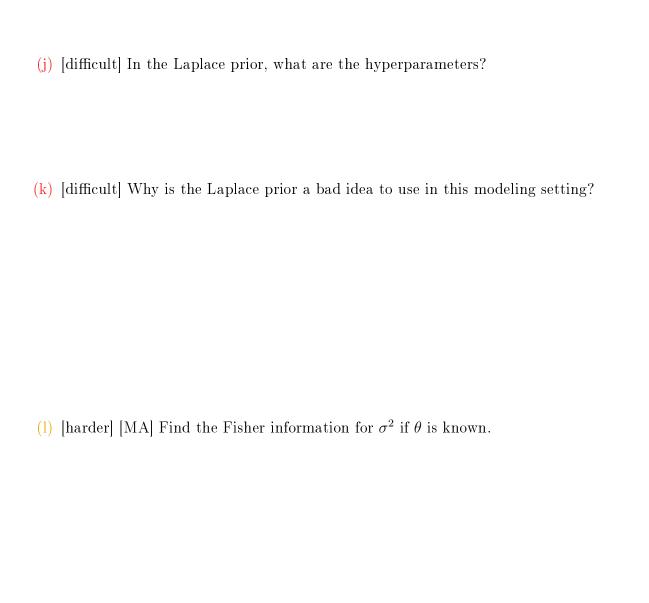This completes the proof that the kernel in (b) is inverse gamma (see lecture 17).

(d) [easy] What is the parameter space of the inverse gamma r.v.? What is its mean? Mode? Is there a formula for the median in closed form?

(e) [easy] Show that posterior of $\sigma^2 \mid X$, θ is inverse gamma (and find the posterior parameters).

(f) [easy] Plot a few PDF's of inverse gammas with different parameters to illustrate the different possible shapes.

(g) [difficult] Show that posterior of $\sigma^2 \mid X,\ \theta$ is inverse gamma (and find the posterior parameters) if $\mathbb{P}\left(\sigma^2\right) = \mathrm{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$. Try to do it yourself and only copy from the notes if you have to.

(h) [easy] What is the pseudodata interpretation of the hyperparameters n_0 and σ_0^2?

(i) [easy] Based on the pseudodata interpretation of the hyperparameters n_0 and σ_0^2, what would Haldane's prior be and why?

(j) [difficult] In the Laplace prior, what are the hyperparameters?

(k) [difficult] Why is the Laplace prior a bad idea to use in this modeling setting?

(l) [harder] [MA] Find the Fisher information for σ^2 if θ is known.

(m) [harder] [MA] Show the Jeffrey's Prior for inference on σ^2 if θ is known is $\propto (\sigma^2)^{-1}$.

(n) [easy] Provide all three Bayesian point estimates for σ^2 given θ.

(o) [easy] Show that the $\hat{\theta}_{\text{MMSE}}$ is a linear shrinkage estimator. Is it valid for every inverse gamma prior?

(p) [harder] Show that the $\hat{\theta}_{\text{MAP}}$ is a linear shrinkage estimator (i.e. a linear combination of the MLE and the prior mode). Is it valid for every inverse gamma prior?

(q) [difficult] [MA] Show that predictive distribution of $X^* \mid X$, θ is non-standard Student's t distribution if $n^* = 1$ and $\theta \sim \text{InvGamma}\left(\frac{n_0}{2}, \frac{n_0 \sigma_0^2}{2}\right)$ by solving the integral. Try to do it yourself and only copy from the notes if you have to!

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)
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)
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 4

This question is about building models for the prices of cars sold at dealerships.

The 2016 Honda Accord sells at many different dealerships in New York City but sell it for more and some for less. We'll assume that the final negotiated price is distributed normally because it's most likely the sum of many different negotiation factors.

Our goal here is to determine the mean price at a certain car dealership in Astoria that people have been saying is "too cheap" and if it's too cheap, Honda corporate may wish to investigate.

(a) [easy] Assume that each Accord's price at the Astoria dealership is normal and $\overset{iid}{\sim}$ given the parameters. Is this a good model? Why or why not? There is no "correct" answer here but I expect you to defend whatever answer you write using the concepts we discussed in class.

(b) [easy] Despite what you wrote in (b), assume $\overset{iid}{\sim}$ for the rest of the problem. The nationwide variance for a Honda Accord selling price we're going to assume is $\sigma^2 = \$1000^2$. Given a sample with average $\bar{x}$ and sample size n, what is the distribution of the mean price of a car from this shady Astoria dealership? Assume an uninformative prior of your choice but ensure to explicitly state it.

(c) [harder] You and your colleague go down to the Astoria dealership undercover and ask to buy a Honda. After much negotiation, they will sell it to you for $19,000 and they will sell it to your colleague for $18,200 but they sense something suspicious so you hesitate to send another one of your guys down there to do another faux negotiation. Unfortunately, we're going to have to estimate the mean with just $x_1 = 19000$ and $x_2 = 18200$. What is your best guess of the mean price of Honda Accords sold here? Assume your prior from (a). Compute explicitly as a number rounded to two decimals.

(d) [easy] What is the shrinkage value (which we have been denoting ρ) for this estimate? Compute explicitly as a number rounded to two decimals.

(e) [harder] Based on this data, we wish to test if this dealership is selling Honda Accords below the manufacturer sugested retail price (MSRP) of $22,205 — if so, they would be subject to a fine. Calculate a p-value for this test below by using notation from Table 1 but do not solve numerically.

(f) [difficult] What is the probability I get a really good deal — that I can buy a car from these Astoria people for under $17,000? Use the notation from Table 1 but do not solve numerically.

(g) [harder] Find a 95% posterior predictive interval (PI) for the price of the next car.

Problem 5

This question is about building a model to understand the accuracy of this beverage-filling machine pictured below:

This machine fills 12oz plastic bottles. There is no doubt the mean amount of liquid filled per bottle is 12oz as been determined by the final weights of pallets of filled bottles. But we are uncertain about the variance. We decide to do an experiment and select $n = 21$ bottles at random and measure the amount of liquid in each bottle. Here are the measurements:

```
12.00 12.05 11.98 11.66 12.05 11.92 12.03
12.23 12.36 11.57 12.04 12.10 11.99 12.47
12.57 11.83 12.20 12.48 12.14 12.14 12.74
```

(a) [easy] Write a parametric model $\mathcal{F}$ below for the amount of milk in each of the n bottles. Hint: use the normal model!

(b) [easy] Find the MLE for σ^2. Compute explicitly as a number rounded to two decimals.

(c) [easy] Under the Jeffrey's prior for σ^2, find the posterior of σ^2 by solving for the parameter values.

(d) [harder] Write an expression for the 95% left-sided credible region for σ^2. (This will give the upper bound for the machine's variance).

(e) [harder] The bottles are actually 13.5oz. This means that you wish to test if $\sigma^2 > 0.352$ for if so, about 1/100,000 of the bottles will be overfull and that's the tolerance of the factory. Write an expression for the Bayesian p-value of this test.

(f) [harder] Write an expression for the probability the next bottle has more than 13oz of liquid.

(g) [harder] Write an expression for the 95% posterior predictive interval (PI) for the amount of liquid in the next bottle.