MATH 341 / 650.3 Spring 2021 Homework #2

Professor Adam Kapelner

Due by email, Sunday 11:59PM, February 28, 2020

(this document last updated Friday 19$^{\text{th}}$ February, 2021 at 11:53am)

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, read about the beta-binomial model and conjugacy in Bolstad and read Ch4–7 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 4-7.

(a) [easy] Describe four things Bayesian modeling was applied to during WWII and identify the people who developed each application.

(b) [harder] What do you think was the main reason Bayesian Statistics fell out of favor at the end of WWII?

(c) [harder] Why weren't the leaders of Statistics world in the 1950's able to answer the think-tank's question about the $\mathbb{P}$ (war in the next 5 years)?

(d) [easy] Who was responsible for reviving the interest in Bayesian Statistics post-WWII and why?

(e) [difficult] In 1955, there were no midair collisions of two planes. How was the actuary able to estimate that the number would be above zero?

(f) [easy] The main attack on Bayesian Statistics has always been subjectivity. Answer the following question how Savage would have answered it: "If prior opinions can differ from one researcher to the next, what happens to scientific objectivity in data analysis?" Do you believe Savage's idea is the way science works in the real world?

(g) [difficult] [MA] On page 104, Sharon writes, "Bayesians would also be able to concentrate on what happened, not on what *could* have happened according to Neyman Pearson's samping plan". (Note that the "Neyman Pearson's samping plan" is synonymous with Frequentist Statistics). Explain (1) how Bayesians concentrate on "what happened" and (2) how Frequentists concentrate on what "*could* have happened" in the context on page 104.

(h) [easy] Who were the two tireless champions of Bayesian Statistics throughout the 50's, 60's and 70's and where geographically were they located during the majority of their career?

Problem 2

We will prove an interesting fact about the posterior being the result of iterative updates. Let $\mathcal{F} : \overset{iid}{\sim} \text{Bernoulli}(\theta)$ and we'll use the parameter space from class that's a subset of the full parameter space, $\Theta = \{0.5, 0.75\}$. We'll also consider the dataset from class where $x_1 = 0, x_2 = 1, x_3 = 1$.

(a) [easy] Using the principle of indifference, what is $\mathbb{P}(\theta = 0.75)$?

(b) [easy] Imagine we saw *only* $x_1 = 0$. What is its likelihood?

(c) [harder] Find the posterior probability of $\theta = 0.75$ when seeing only $x_1 = 0$.

(d) [easy] Find the posterior probability of $\theta = 0.5$ when seeing only $x_1 = 0$.

(e) [harder] Use the result from (c, d) as the prior distribution now. This is strange but go with it. It is no longer 50-50 as it should be tilted more towards $\theta = 0.5$ then $\theta = 0.75$ since the answer for (d) should be greater than the answer for (c). Using this prior, calculate the posterior for both values of θ for only $x_2 = 1$ (we are ignoring x_1 explicitly since it's implicitly factored into the prior).

(f) [harder] Third time is a charn. Use the result from (e) as the prior distribution now. Using this prior, calculate the posterior for only $x_3 = 1$ (we are ignoring x_1 and x_2 explicitly since they're implicitly factored into the prior). Your answer should match the posterior from class (probability 0.53 for $\theta = 0.75$ and probability 0.47 for $\theta = 0.5$) where we considered all x_1, x_2, x_3 in one shot together.

(g) [difficult] [MA] Bayesian inference can be thought of as "begin with prior, see one data point, update distribution of θ, if you see another data point, use that updated distribution as prior to then compute a second updated distributed, etc. etc." Via induction, prove this idea under the case where $\mathcal{F}$ is a general $\overset{iid}{\sim}$ rv model. The base case is done (it's Bayes rule for X_1 and θ). Then assume the theorem for $n-1$ and prove the inductive step:

$$\mathbb{P}(\theta \mid X_1, \ldots, X_n) = \frac{\mathbb{P}(X_n \mid \theta)\overbrace{\mathbb{P}(\theta \mid X_1, \ldots, X_{n-1})}^{\substack{\text{this prior is assumed} \\ \text{to be the posterior for the} \\ \text{previous } n-1 \text{ data points}}}}{\mathbb{P}(X_1, \ldots, X_{n-1})}$$

Problem 3

We will now be looking at the beta-prior for $\mathcal{F} : \overset{iid}{\sim} \text{Bernoulli}(\theta)$ and we'll use the full parameter space, $\Theta = (0, 1)$.

(a) [easy] Using the principle of indifference, what should the prior on θ (the parameter for the Bernoulli model) be?

(b) [harder] [MA] Can any discrete distribution satisfy the principle of indifference? Prove or disprove.

(c) [easy] Let's say $n = 6$ and your data is $0, 1, 1, 1, 1, 1$. What is the likelihood of this event?

(d) [easy] Does it matter the order as to which the data came in? Yes/no.

(e) [harder] Show that the unconditional probability (the prior predictive distribution, the denominator in Bayes rule) is a beta function and specify its two arguments.

(f) [harder] Calculate this beta function explicitly.

(g) [harder] Put your answers together to find the posterior probability of θ given this dataset. Do not use the beta function in your answer. Plot this posterior density function as best as you can.

(h) [easy] Sketch / plot / illustrate this posterior density function as best as you can by hand. Indicate where the $\hat{\theta}_{\text{MAP}}$, $\hat{\theta}_{\text{MMSE}}$ and $\hat{\theta}_{\text{MMAE}}$ estimates for θ are on the illustration using vertical lines as best as you can. No need to calculate them now explicitly. Just eyeball it on your drawing.

(i) [harder] Show that the posterior is a beta distribution and specify its parameters.

(j) [difficult] [MA] Prove that the posterior expectation is the optimal estimator given your prior and the data under mean squared error loss.

(k) [E.C.] Prove that the posterior median is the optimal estimator given your prior and the data under mean absolute error loss.

(l) [easy] Now imagine you are not indifferent and you have some idea about what θ could be a priori and that subjective feeling can be specified as a beta distribution. (1) Draw the basic shapes that the beta distribution can take on, (2) give an example of α and β values that would produce these shapes and (3) write a sentence about what each one means for your prior belief. These shapes are in the notes.

(m) [harder] Imagine n data points of which you don't know the specific realization values but you know their sum denoted x, the total number of 1's and thus $n - x$ is the total number of 0's. This is now the case where $\mathcal{F}$: one realization of a Binomial (n, θ) where n is known but θ is unknown.

Using your prior of $\theta \sim \text{Beta}(\alpha, \beta)$, show that $\theta \mid X \sim \text{Beta}(\alpha + x, \beta + (n - x))$.

(n) [easy] What does it mean that the beta distribution is the "conjugate prior" for the binomial likelihood?

(o) [harder] [MA] Show that if $Y \sim \text{Beta}(\alpha, \beta)$ then $\mathbb{V}\text{ar}[Y] = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

(p) [harder] The posterior is $\theta \mid X \sim \text{Beta}(\alpha + x, \beta + (n - x))$. Some say the values of α and β can be interpreted as follows: α is considered the prior number of successes and β is considered the prior number of failures. Why is this a good interpretation? Writing out the PDF of $\theta \mid X$ should help you see it.

(q) [harder] If you employ the principle of indifference, how many successes and failures is that equivalent to seeing a priori?

(r) [easy] Why are large values of α and/or β considered to compose a "strong" prior?

(s) [harder] [MA] What is the "weakest" prior you can think of and why?

(t) [difficult] I think a priori that θ should be expected to be 0.8 with a standard error of 0.02. Solve for the values of α and β based on my a priori specification.

(u) [easy] Assume the dataset in (b) where $n = 6$. Assume $\theta \sim \text{Beta}\,(\alpha = 2,\, \beta = 2)$ a priori. Find the $\hat{\theta}_{\text{MAP}}$, $\hat{\theta}_{\text{MMSE}}$ and $\hat{\theta}_{\text{MMAE}}$ estimates for θ. For the $\hat{\theta}_{\text{MMAE}}$ estimate, you'll need to obtain a quantile of the beta distribution. Use `R` on your computer or online using rextester. The `qbeta` function in `R` finds arbitrary beta quantiles. Its first argument is the quantile desired e.g. 2.5%, the next is α and the third is β. So to find the 97.5%ile of a $\text{Beta}\,(\alpha = 2,\, \beta = 2)$ for example you type `qbeta(.975, 2, 2)` into the `R` console.

(v) [harder] Why are all three of these estimates the same?