MATH 341 / 650.3 Spring 2018 Homework #2

Professor Adam Kapelner

Due *in class*, Friday 5PM, March 2, 2018

(this document last updated Wednesday 21ˢᵗ February, 2018 at 12:23am)

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 Math 241 concerning random variables, support, parameter space, expectation, variance, mode, quantiles, uniform distribution, binomial distribution, beta distribution, conjugacy, bayesian credible sets, 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 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 $\mathbb{P}(\text{war in the next 5 years})$?

(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 now be looking at the beta-prior, binomial-likelihood Bayesian model.

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

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

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

(d) [harder] Show that the unconditional joint probability (the denominator in Bayes rule) is a beta function and specify its two arguments.

(e) [harder] Put your answer from (a), (b) and (d) together to find the posterior probability of θ given this dataset. Show that it is equal to a beta distribution and specify its parameters.

(f) [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.

(g) [harder] Imagine n data points of which you don't know the realization values. Using your prior of $\theta \sim \text{Beta}\,(\alpha,\,\beta)$, show that $\theta \mid X \sim \text{Beta}\,(\alpha + x,\,\beta + (n - x))$. Note that $x := \sum_{i=1}^{n} x_i$ which is the total number of successes and thereby $n - x$ is the total number of failures.

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

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

(j) [E.C.] Prove that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$.

(k) [harder] Stare at that distribution, $\theta \mid X \sim \text{Beta}(\alpha + x, \beta + (n - x))$. 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.

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

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

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

(o) [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.

(p) [difficult] Prove that the posterior predictive distribution is $X^* \mid X \sim \text{Bernoulli}\left(\frac{x+\alpha}{n+\alpha+\beta}\right)$. MA students — do this yourself. Other students — use my notes and justify each step. I use a property of the gamma function. Remember, if $W \sim \text{Bernoulli}(\theta)$ then $\mathbb{P}(W=1) = \theta$. Use that trick! Set $X^* = 1$ and find that probability!

(q) [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 $\hat{\theta}_{\text{MAE}}$ estimates for θ. For the $\hat{\theta}_{\text{MAE}}$ 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.

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

(s) [easy] Write out an expression for the 95% credible region for θ. Then solve computationally using the `qbeta` function in `R`.

(t) [easy] Compute a 95% frequentist CI for θ.

(u) [difficult] Let $\mu : \mathbb{R} \to \mathbb{R}^+$ be the Lebesgue measure which measures the length of a subset of $\mathbb{R}$. Why is $\mu(\mathrm{CR}) < \mu(\mathrm{CI})$? That is, why is the Bayesian Confidence Interval tighter than the Frequentist Confidence Interval? Use your previous answers.

(v) [easy] Explain the disadvantages of the highest density region method for computing credible regions.

(w) [harder] Design a prior where you believe $\mathbb{E}[\theta] = 0.5$ and you feel as if your belief represents information contained in five coin flips.

(x) [harder] Calculate a 95% a priori credible region for θ. Use R on your computer (or R-Fiddle online) and its `qbeta` function.

(y) [easy] You flip the same coin 100 times and you observe 39 heads. Find the distribution of $\theta \mid X$.

(z) [easy] Calculate a 95% a posteriori credible region for θ.

(aa) [easy] Why is your answer to (d) a smaller interval than (b)?