MATH 341 / 650 Spring 2017 Homework #3

Professor Adam Kapelner

Due 2PM under my office door (KY604), Monday, March 6, 2017

(this document last updated Sunday 26$^{\text{th}}$ February, 2017 at 7:24pm)

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 Binomial-Beta, then read about Bayesian Hypothesis Testing, Bayes Factors, Credible Regions and Empirical Bayes.

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

We will now be looking at the beta-prior, binomial-likelihood Bayesian model once again.

(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. We did this in class.

(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 five 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. Show that $\theta \mid X \sim \text{Beta}\left(\alpha + x,\, \beta + (n - x)\right)$. Note that $x := \sum_{i=1}^{n} x_i$ which is the total number of successes and thereby $n - x$ is the total nubmer of failures. The answer is in the notes but try to do it yourself.

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

(i) [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: $\alpha - 1$ is considered the prior number of successes and $\beta - 1$ 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.

(j) [harder] By the principle of indifference, how many successes and failures is that equivalent to seeing a priori?

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

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

(m) [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.

(n) [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.

(o) [harder] The frequentist estimate of θ is $\hat{p} = 3/6 = 0.5$ So a frequentist would probably use a posterior predictive distribution (if he had such a thing) as $X^* \mid X \sim \text{Bernoulli}(0.5)$. Why conceptually does this answer differ from your answer in (n)?

(p) [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 R-Fiddle. 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.

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

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

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

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

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

Problem 2

This problem is concerned with the logical definition of probability. As a review, we have:

1. **The Objective View** — This is the view that probabilities are properties of the physical world and can only be defined by either

(a) its **Long Run Frequency** Seeing the same event over and over again and tabulating when the event occurs will create a frequency which will asympotically become the probability or

(b) its **Propensity** which means deep down inside, the physical object is wired for events in certain proportions.

Thus, events that are non physical such as the probability of Donald Trump winning the 2016 election is outside of the purview of probability. The objective view of probability is tied to the frequentist view of statistics. We also have the...

2. **The Epistemic View** — This is the view that probabilities are inherently living inside the minds of human beings who are forced to grapple with uncertainty as they see it. Laplace believed probability is an illusion because we don't have certainty about the universe. The two definitions here are that probability...

   (a) is **Logical** which means that given the same information, everyone would come to the same conclusion.

   (b) is **Subjective** which means that given the same information, everyone would *not* come to the same conclusion. Thus probability is defined as the degree of belief of some individual which differs from another individual.

   Thus events that are non physical such as the probability of Donald Trump winning the 2016 election (prior to the election) can now be legitimate probabilities as they can be computed.

(a) [easy] We discussed last time that the logical definition requires the principle of indifference which goes by many different names. We will now go about showing that the principle of indifference has a tenuous foundation and thereby rendering the logical theory of probability inadequate. Thus, our conclusion will be that Bayesian Statistics runs on the Subjective definition which we may develop in a later homework. We begin with demonstrating a paradox in the logical definition for a discrete set of θ_0.

   Imagine you have a library with thousands of books but all are either red, green, yellow or purple but you don't know the proportions of the books' colors. Imagine you are blindfolded and select a random book and you are only interested if it's *red* or *not red*. According to the principle of indifference, what is your prior probability that the book is red? Remember, $|\Theta_0| = 2$ here.

(b) [easy] Imagine you are blindfolded and select a random book and you are interested if it's red, green, yellow or purple. According to the principle of indifference, what is your prior probability that the book is red? Remember, $|\Theta_0| = 4$ here.

(c) [harder] Why do (a) and (b) constitute a paradox? Does this limit the application of the principle of indifference?