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 9:29pm)

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

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

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

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

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

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

(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. Remember, if $W \sim \text{Bernoulli}(\theta)$ then $\mathbb{P}(W = 1) = \theta$. Use that trick! Set $X^* = 1$ and find that probability!

(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

We will again be looking at the beta-prior, binomial-likelihood Bayesian model and practicing hypothesis testing.

(a) [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.

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

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

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

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

(f) [harder] Test the hypothesis that this coin is fair given prior information from (a) and the data from (c). Use the credible region method. Make sure you say whether you retain or reject the null and justify why.

(g) [harder] Test the hypothesis that this coin has a bias towards Heads (not tails) given prior information from (a) and the data from (c). Calculate the Bayesian p-val for the test to determine if you should retain or reject H_0.

(h) [easy] Let's say you wanted to test whether the coin is fair but you are indifferent to any θ which is different from 0.5 by a margin of 0.1. Write out the hypotheses for this test.

(i) [harder] Test the hypotheses from (h) given prior information from (a) and the data from (c). Make sure you say whether you retain or reject the null and justify why.

(j) [harder] Calculate the Bayesian p-val for the test in (i).

(k) [harder] Given the tested hypotheses from (i) and data from (c), write the formula for the Bayes factor and then write the integral expression. Do not solve.

(l) [harder] [MA] Calculate the Bayes Factor numerically by solving the integral expression in (k). Interpret your value of K (or B) according to wikipedia page about Bayes Factors.

(m) [easy] Given prior information from (a) and the data from (c), what is the distribution of one future coin flip?

(n) [harder] Prove that the $\hat{\theta}_{\text{MMSE}}$ is a shrinkage estimator.

(o) [easy] Assume again the prior information from (a). What is the shrinkage proportion ρ for this prior when estimating θ via $\hat{\theta}_{\text{MMSE}}$?.

(p) [difficult] Prove that $\hat{\theta}_{\text{MMSE}}$ is a biased estimator (i.e. its expectation is *not* θ).

(q) [easy] Prove that $\lim_{n \to \infty} \rho = 0$ and therefore this bias $\to 0$ as your dataset gets larger.

(r) [difficult] [MA] Why on Earth should anyone use shrinkage estimators if they're biased? Google it. Discuss.