MATH 341 / 650.3 Spring 2020 Homework #2

Frank Palma Gomez

Sunday 16^{th} February, 2020

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 now be looking at the beta-prior, binomial-likelihood Bayesian model and introduce credible regions as well.

(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 joint probability (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 mean squared error optimal estimator given your prior and the data.

(k) [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.

(l) [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.

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

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

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

(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?

(w) [easy] Write out an expression for the 95% credible region for θ. Then write out the answer using the `qbeta` function from the `R` language.

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

(y) [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 previous answers.

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

(aa) [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.

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

(cc) [easy] You flip the same coin 100 times and you observe 39 heads. Calculate a 95% a posteriori credible region for θ. Round to the nearest 3 decimal points.