

MATH 341 / 650 Spring 2017 Homework #4

Professor Adam Kapelner

Due 2PM under my office door (KY604), Monday, March 20, 2017

(this document last updated Wednesday 8th March, 2017 at 4:17pm)

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 L^AT_EX. Links to installing L^AT_EX and program for compiling L^AT_EX 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 L^AT_EX, 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 sampling plan”. (Note that the “Neyman Pearson’s sampling 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

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 asymptotically 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 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?
- (d) [easy] Now we're going to work on trashing the principle of indifference for continuous measures. Thanks to Wikipedia... Imagine I have a cube-shaped box. The length of the side we call S and we know it's less than 1 inch in length. Thus its prior distribution under the principle of indifference should be $S \sim U(0, 1)$. What is the expected length of the side a priori (that is according to the prior distribution)? Please do not overthink this.
- (e) [harder] Given the same prior as previously, imagine the surface area which is calculated as $6S^2$. Find expectation of the surface area. You may need to look at your Math 241 notes to get the variance of the uniform r.v. Is the answer you get equal to using the prior mean length and computing the surface area based on that, i.e. $6\mathbb{E}[S]^2$?

(f) [harder] We will not prove this, but it shouldn't come as a surprise that the surface area is not distributed uniformly between 0 and 6 given what you saw in your last answer. What does this mean for the principle of indifference? It only works...

(g) [easy] It may be argued that one is only indifferent to the length of the side and that it's okay this implies a non-indifference to the surface area and volume. But here's a paradox that's harder to argue with which I got from Gillies' book. Imagine you have a drink in front of you made up of some parts wine and some parts water. You have a prior belief that the ratio of wine/water is $U(1/4, 4)$ using the principle of indifference. What is the probability (according to your prior belief) that the $\mathbb{P}(\text{wine/water} \geq 2)$?

(h) [easy] Using the principle of indifference, what does this imply that the ratio of water/wine should be $U(1/4, 4)$ as well? It shouldn't matter whether you pick wine divided by water or water divided by wine, right? Answer yes/no and give your gut reaction.

(i) [easy] Assuming that the ratio of water/wine is $U(1/4, 4)$, calculate:

$$\mathbb{P}((\text{wine/water})^{-1} \geq 2^{-1}) = \mathbb{P}\left(\text{water/wine} \leq \frac{1}{2}\right) =$$

And: is this different to the answer you got in (g)?

(j) [difficult] Why is this a paradox? Write a couple sentences how this should doom the principle of indifference in the case of a continuous Θ_0 space.

(k) [harder] [MA] Now we're going to see how this fails. Under the prior belief that the ratio of wine/water is $U(1/4, 4)$, derive the PDF of the ratio of water/wine. Is it also $U(1/4, 4)$? If you need a refresher on this stuff, see here. (Note: we're going to see Jeffrey's answer to this issue soon enough in class).

Problem 3

Some quick question on mixture distributions.

(a) [easy] If X is independent to W and X is independent to Z and X is independent to U , can you write $\mathbb{P}(X \mid U, V, W, Y, Z)$ more compactly? Do so below.

(b) [easy] Let X be $\mathcal{N}(0, 1^2)$ $1/3$ of the time and $\text{Exp}(3)$ $2/3$ of the time. What is its pdf?

- (c) [difficult] Let's say $X \mid \beta \sim \text{Beta}(1, \beta)$ where $\beta \mid \lambda \sim \text{Exp}(\lambda)$. Write an integral expression which when solved, finds the compound / marginal density of X . DO NOT solve.

- (d) [difficult] [MA] Let's say $X \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ where $\theta \mid \mu_0, \tau^2 \sim \mathcal{N}(\mu_0, \tau^2)$. Write an integral expression which when solved, finds the compound / marginal density of X . DO NOT solve.

Problem 4

We will now have lots of examples finding kernels from common distributions. Some of these questions are silly, but they will force you to think hard about what the kernel is under different situations. And... they're fun!

- (a) [easy] What is the kernel of $X \mid \theta, n \sim \text{Binomial}(n, \theta)$?

(b) [difficult] What is the kernel of $X, n \mid \theta \sim \text{Binomial}(n, \theta)$? Be careful...

(c) [easy] What is the kernel of $X \mid \theta \sim \text{Poisson}(\theta)$?

(d) [difficult] What is the kernel of $\theta \mid X \sim \text{Poisson}(\theta)$? Be careful...

(e) [easy] What is the kernel of $X \mid \alpha, \beta \sim \text{Beta}(\alpha, \beta)$?

(f) [easy] What is the kernel of $X \mid \theta \sim \text{Exp}(\theta)$?

(g) [easy] What is the kernel of $X \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$?

(h) [difficult] What is the kernel of $\theta, \sigma^2 \mid X \sim \mathcal{N}(\theta, \sigma^2)$? Be careful...

(i) [harder] [MA] What is the kernel of

$$X \mid N, \theta, n \sim \text{Hypergeometric}(N, \theta, n) := \frac{\binom{\theta}{x} \binom{N-\theta}{n-x}}{\binom{N}{n}}$$

where N is the number of total balls in the bag, θ is the number of success balls in the bag and n is the number drawn out of the bag?

- (j) [difficult] [MA] If $X \mid \theta, \sigma^2 \sim \mathcal{N}(\theta, \sigma^2)$ and $\theta \mid \mu_0, \tau^2 \sim \mathcal{N}(\mu_0, \tau^2)$, what is the kernel of $\theta \mid X, \sigma^2, \mu_0, \tau^2$?

Problem 5

These are questions about other vague priors: improper priors and Jeffreys priors.

- (a) [easy] What is an improper prior?
- (b) [easy] Is $\theta \sim \text{Beta}(100, 0)$ improper? Yes / no and provide a proof.
- (c) [easy] When are improper priors “legal”?

(d) [easy] When are improper priors “illegal”?

(e) [difficult] What does $I(\theta)$ tell you about the random variable with respect to its parameter θ ?

(f) [harder] If I compute a posterior on the θ scale and then measure the parameter on another scale, will I (generally) get the same posterior probability? Yes/no explain.

(g) [easy] What is the Jeffrey’s prior for θ under the binomial likelihood? Your answer must be a distribution.

(h) [difficult] What is the Jeffrey's prior for $\theta = t^{-1}(r) = \frac{e^r}{1+e^r}$ (i.e. the log-odds reparameterization) under the binomial likelihood?

(i) [difficult] Explain the advantage of Jeffrey's prior.

- (j) [difficult] [MA] Prove Jeffrey's invariance principle i.e. prove that the Jeffrey's prior makes your prior probability immune to transformations. Use the second proof from class.

Problem 6

This question is about “batting averages” in baseball.



Every hitter's *sample* batting average (BA) is defined as:

$$BA := \frac{\text{sample \# of hits}}{\text{sample \# of at bats}}$$

In this problem we care about estimating a hitter's *true* batting average which we call θ . Each player has a different θ but we focus in this problem on one specific player. In order to estimate the player's true batting average, we make use of the sample batting average as defined above (with Bayesian modifications, of course).

We assume that each at bat (for any player) are conditionally $\overset{iid}{\sim}$ based on the players' true batting average, θ . So if a player has n at bats, then each successful hit in each at bat can be modeled via $X_1 | \theta, X_2 | \theta, \dots, X_n | \theta \overset{iid}{\sim} \text{Bernoulli}(\theta)$ i.e. the standard assumption.

- (a) [easy] Looking at the entire dataset for 6,061 batters who had 100 or more at bats, I fit a beta function to the sample batting averages and estimated $\alpha = 42.3$ and $\beta = 127.7$ (which we called “empirical Bayes” estimates in class). Consider building a prior from this estimate as $\theta \sim \text{Beta}(42.3, 127.7)$ Would a prior based on these hyperparameter estimates be “objective”? Yes / No. Why?
- (b) [easy] Using prior data to build the prior is called...
- (c) [easy] Is the prior from (a) considered a “conjugate prior”? Yes / No.
- (d) [easy] Using the prior from (a), find the $\hat{\theta}_{\text{MMSE}}$.
- (e) [difficult] We now observe four at bats for a new player and there were no hits. Write an exact expression for the batter getting 14 or more hits on the next 20 at bats. You can leave your answer in terms of the beta function. Do not compute explicitly.