

# MATH 390 / 650.3 Spring 2022 Homework #4

Professor Adam Kapelner

Due by email 11:59PM April 5, 2022

(this document last updated Sunday 27<sup>th</sup> March, 2022 at 9:06pm)

## 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, read about Jeffrey’s objective prior, informative priors / empirical Bayes, the Poisson-Gamma conjugate model and read chapters 8-11 in 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 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 L<sup>A</sup>T<sub>E</sub>X. Links to installing L<sup>A</sup>T<sub>E</sub>X and program for compiling L<sup>A</sup>T<sub>E</sub>X is found on the syllabus. You are encouraged to use [overleaf.com](https://www.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<sup>A</sup>T<sub>E</sub>X, 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 11–14.

- (a) [easy] Did Savage like Schlaifer? Yes / No and why?
  
  
  
  
  
  
  
  
  
  
- (b) [easy] How did Neyman-Pearson approach statistical decision theory? What is the weakness to this approach? (p145)
  
  
  
  
  
  
  
  
  
  
- (c) [easy] Who popularized “probability trees” (and “tree flipping”) similar to exercises we did in Math 241?
  
  
  
  
  
  
  
  
  
  
- (d) [easy] Where are Bayesian methods taught more widely than any other discipline in academia?
  
  
  
  
  
  
  
  
  
  
- (e) [easy] Despite the popularity of his Bayesian textbook on business decision theory, why didn't Schlaifer's Bayesianism catch on in the real world of business executives making decisions?

(f) [easy] Why did the pollsters fail (big time) to predict Harry Truman's victory in the 1948 presidential election?

(g) [easy] When does the difference between Bayesianism and Frequentism grow "immense"?

(h) [easy] How did Mosteller demonstrate that Madison wrote the 12 Federalist papers of unknown authorship?

(i) [easy] Write a one paragraph biography of John Tukey.

(j) [easy] Why did Alfred Kinsey's wife want to poison John Tukey?

(k) [easy] Tukey helped NBC with polling predictions for the presidential campaign. What was NBC's polling algorithm based on?

(l) [easy] Why is “objectivity an heirloom ... and ... a fallacy?”

(m) [easy] Why do you think Tukey called Bayes Rule by the name “borrowing strength?”

(n) [easy] Why is it that we don't know a lot of Bayes Rule's modern history?

(o) [easy] Generally speaking, how does Nate Silver predict elections?

(p) [easy] How many Bayesians of import were there in 1979?

(q) [easy] What advice did Chernoff give to Susan Holmes? (Note: Susan Holmes was my undergraduate advisor).

(r) [easy] How did Rasmussen's team estimate the probability of a nuclear plant core meltdown?

(s) [easy] How did the Three Mile Island accident vindicate Rasmussen's committee report?

## **Problem 2**

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

(a) [easy] What is an improper prior?

(b) [easy] When are improper priors "legal"?

(c) [easy] When are improper priors "illegal"?

- (d) [difficult] [MA] What does  $I(\theta)$  tell you about the random variable with respect to its parameter  $\theta$ ?
- (e) [harder] If I compute a posterior on the  $\theta$  scale and then measure the parameter on another scale, will I (generally) get the same posterior probability? Yes/no explain.
- (f) [easy] What is the Jeffrey's prior for  $\theta$  under the binomial likelihood? Your answer must be a distribution.
- (g) [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?

(h) [difficult] Explain the advantage of Jeffrey's prior in words. Feel free to use Google.

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

### Problem 3

This question is about estimation of “true career 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* career batting average which we call  $\theta$ . Each player has a different  $\theta$  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,  $\theta$ . 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 and thus the total number of hits out of  $n$  at bats is binomial.

Looking at the entire dataset for 6,061 batters who had 100 or more at bats, I fit the beta distribution PDF to the sample batting averages using the maximum likelihood approach and I’m estimating  $\alpha = 42.3$  and  $\beta = 127.7$ . Consider building a prior from this estimate as  $\theta \sim \text{Beta}(42.3, 127.7)$

- (a) [easy] Is the prior “conjugate”? Yes / No.
- (b) [easy] Is this prior “indifferent”? Yes / No.
- (c) [easy] Is this prior “objective”? Yes / No.
- (d) [easy] Is this prior “informative”? Yes / No.
- (e) [easy] Using prior data to build the prior is called...
- (f) [easy] This prior has the information contained in how many observations?



- (g) [easy] We now observe four at bats for a new player and there were no hits. Find the  $\hat{\theta}_{\text{MMSE}}$ .
- (h) [easy] Why was your answer so far away from  $\hat{\theta}_{\text{MLE}} = 0$ ? What is the shrinkage proportion in this estimation?
- (i) [harder] Why is it a good idea to shrink so hard here? Why do some consider this to be one of the beauties of Bayesian modeling?
- (j) [difficult] 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.
- (k) [harder] How many hits is the batter expected to get in the next 20 at bats?

## Problem 4

We will ask some basic problems on the Gamma-Poisson conjugate model.

- (a) [easy] Write the PDF of  $\theta \sim \text{Gamma}(\alpha, \beta)$  which is the gamma distribution with the standard parameterization and notated with the hyperparameters we used in class.
- (b) [easy] What is the support and parameter space?
- (c) [easy] What is the expectation and standard error and mode?
- (d) [easy] Draw four different pictures of different hyperparameter combinations to demonstrate this model's flexibility
- (e) [harder] Prove that the Poisson likelihood for  $n$  observations, i.e.  $X_1, \dots, X_n; \theta \stackrel{iid}{\sim} \text{Poisson}(\theta)$ , with a gamma prior yields a gamma posterior and find its parameters.

- (f) [easy] Now that you see the posterior, provide a pseudodata interpretation for both hyperparameters.
- (g) [harder] Find the Bayesian point estimates as function of the data and prior's hyperparameters (i.e.  $\hat{\theta}_{\text{MMSE}}$ ,  $\hat{\theta}_{\text{MMAE}}$  and  $\hat{\theta}_{\text{MAP}}$ ).
- (h) [harder] If  $X_1, \dots, X_n; \theta \stackrel{iid}{\sim} \text{Poisson}(\theta)$ , find  $\hat{\theta}_{\text{MLE}}$ .
- (i) [harder] Demonstrate that  $\hat{\theta}_{\text{MMSE}}$  is a shrinkage estimator and find  $\rho$ .

- (j) [harder] Demonstrate that  $\mathbb{P}(\theta) \propto 1$  is improper.
- (k) [easy] [MA] Demonstrate that  $\mathbb{P}(\theta) \propto 1$  can be created by using an improper Gamma distribution (i.e. a Gamma distribution with parameters that are not technically in its parameter space and thereby does not admit a distribution function).
- (l) [harder] Find Jeffrey's prior for the Poisson likelihood model. Try to do it yourself.
- (m) [easy] What is the equivalent of the Haldane prior in the Binomial likelihood model for the Poisson likelihood model? Use an interpretation of pseudocounts to explain.

- (n) [difficult] If  $\mathbb{P}(\theta) = \text{Gamma}(\alpha, \beta)$  where  $\alpha \in \mathbb{N}$ , prove that prior predictive distribution is  $\mathbb{P}(X) = \text{NegBin}(r, p) := \binom{x+r-1}{r-1} (1-p)^x p^r$  where  $p = \beta/(\beta+1)$  and  $r = \alpha$ . This is a little bit different than that posterior predictive distribution derivation we did in class but mostly the same.

- (o) [harder] Why is the extended negative binomial r.v. also known as the gamma-Poisson mixture distribution? Why is it also called the “overdispersed Poisson”?

- (p) [harder] If you observe 0, 3, 2, 4, 2, 6, 1, 0, 5, give a 90% CR for  $\theta$ . Pick a principled objective (uninformative) prior.

(q) [harder] Using the data and the prior from (p), test if  $\theta < 2$ .

(r) [difficult] Using the data and the prior from (p), find the probability the next observation will be a 7. Leave in exact form using Table 1's notation.

(s) [easy] Use the R calculator (if you don't have it on your computer, go to `rdrr.io`) to compute the probability from the last problem to the nearest two significant digits.

(t) [E.C.] [MA] Find the joint posterior predictive distribution for  $m$  future observations. I couldn't find the answer to this myself nor compute the integral.