MATH 341/641 Fall 2023 Homework #6

Professor Adam Kapelner

Due by email 11:59PM Dec 4, 2023

(this document last updated Friday 24$^{\text{th}}$ November, 2023 at 2:49pm)

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 MATH 340 concepts: the Bernoulli, Binomial, Beta and kernels.

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. "[MA]" are for those registered for 621 and extra credit otherwise.

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 5 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 prove an interesting fact about the posterior being the result of iterative updates. Let the DGP be $\overset{iid}{\sim}$ Bernoulli (θ) and we'll use the parameter space from class that's a subset of the full parameter space, $\Theta = \{0.5, 0.75\}$.

(a) [easy] Using the principle of indifference, what is $\mathbb{P}(\theta = 0.75)$?

(b) [harder] Find the posterior probability of $\theta = 0.75$ when seeing only $x_1 = 0$.

(c) [harder] Use the result from (c) as the prior distribution now i.e. let the prior be $\mathbb{P}(\theta \mid X_1 = 0)$. Using this prior, calculate the posterior i.e. $\mathbb{P}(\theta \mid X_2)$

Problem 2

We will now be looking at the beta prior for the $\overset{iid}{\sim}$ Bernoulli (θ) DGP and we'll provide Bayesian Inference on the full parameter space, $\Theta = (0, 1)$.

(a) [easy] Using the principle of indifference, what is the prior the parameter for the Bernoulli model $f(\theta)$?

(b) [easy] Let's say $n = 6$ and your data is $0, 1, 1, 1, 1, 1$. What is the likelihood $\mathbb{P}(\boldsymbol{X} \mid \theta)$ of this event as a function of θ?

(c) [easy] Does it matter the order as to which the data came in? Yes/no.

(d) [harder] Show that the unconditional probability (the prior predictive distribution, the denominator in Bayes rule) is a beta function and specify its two arguments.

(e) [easy] Calculate this beta function as a real value.

(f) [harder] Put the three answers above $f(\theta)$, $\mathbb{P}(\boldsymbol{X} \mid \theta)$, $\mathbb{P}(\boldsymbol{X})$, together to find the posterior probability $f(\theta \mid \boldsymbol{X})$. Do not use the beta function in your answer.

(g) [easy] Show that the posterior is a beta distribution and specify its parameters.

(h) [easy] Calculuate $\hat{\hat{\theta}}^{\text{MAP}}$ and $\hat{\hat{\theta}}^{\text{MMSE}}$ exactly from formulas. Approximate $\hat{\hat{\theta}}^{\text{MMAE}}$ via the `qbeta` function by using `R` on your computer (or use rdrr.io online).

(i) [easy] Compute a 95% frequentist confidence interval (CI) for θ. Is there a problem with it? If so what is wrong and why is it a problem?

(j) [easy] Create a 95% credible region for θ. Use the `qbeta` function and `R` on your computer (or use rdrr.io online).

(k) [easy] Sketch / plot / illustrate this posterior density function as best as you can by hand. Mark $\hat{\hat{\theta}}^{\text{MAP}}$, $\hat{\hat{\theta}}^{\text{MMSE}}$ and $\hat{\hat{\theta}}^{\text{MMAE}}$ and mark them in the drawing as dotted vertical lines. Also, mark the 95% CR for θ underneath the graph.

(l) [easy] Test $H_a : \theta < 0.5$ at $\alpha_0 = 5\%$. Calculate the Bayesian p-value via the `pbeta` function by using `R` on your computer (or use rdrr.io online).

(m) [easy] Test $H_a : \theta > 0.5$ at $\alpha_0 = 5\%$. Calculate the Bayesian p-value via the `pbeta` function by using `R` on your computer (or use rdrr.io online).

(n) [harder] Test $H_a : \theta \neq 0.5$ at $\alpha_0 = 5\%$ and $\delta = 2\%$. Calculate the Bayesian p-value via the `pbeta` function by using `R` on your computer (or use rdrr.io online). Use

(o) [easy] Test $H_a : \theta \neq 0.5$ at $\alpha_0 = 5\%$ by using the 2-sided CR procedure (you already calculated $CR_{\theta,95\%}$ so just check is θ_0 is outside the interval to reject or not).

(p) [difficult] [MA] Find the Bayesian p-value ujsing the CR procedure.

(q) [harder] Let the DGP now be $X \sim \text{Binomial}\,(n,\, \theta)$ where n is known but θ is unknown. Using your prior of $\theta \sim \text{Beta}\,(\alpha,\, \beta)$, show that $\theta \mid X \sim \text{Beta}\,(\alpha + x,\, \beta + (n - x))$. Best to use kernels!

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

(s) [easy] What is Haldane's prior $\mathbb{P}(\theta)$ in the beta-binomial Bayesian model? What was he trying to accomplish with this prior?

(t) [easy] What is the definition of "a proper prior"? Is Haldane's prior $\mathbb{P}(\theta)$ a proper prior? Yes/no and why.

(u) [easy] How many pseudotrials n_0, pseudosuccesses x_0 and pseudofailures $n_0 - x_0$ are contributed by Haldane's prior? Is it the same as for Laplace's prior?

(v) [easy] What is $\hat{\theta}^{\text{MMSE}}$ under Haldane's prior? Is it the same as $\hat{\theta}_{\text{MLE}}$?

(w) [harder] If you employ the principle of indifference, how many successes and failures is that equivalent to seeing a priori?

(x) [harder] What is the "weakest" proper beta prior you can think of and why?

(y) [difficult] Bayesian inference can be thought of as "begin with prior, see one data point, update distribution of θ, if you see another data point, use that updated distribution as prior to then compute a second updated distributed, etc. etc." Prove this idea under the case where the DGP is a general $\overset{iid}{\sim}$ rv model. This amounts to proving for any n that

$$\mathbb{P}\left(\theta \mid X_n\right) = \frac{\mathbb{P}\left(X_n \mid \theta\right) \overbrace{\mathbb{P}\left(\theta \mid X_1, \ldots, X_{n-1}\right)}^{\substack{\text{this prior is assumed} \\ \text{to be the posterior for the} \\ \text{previous } n-1 \text{ data points}}}}{\mathbb{P}\left(X_n\right)}$$

The hard part of this is understanding what the denominator means in this context.

Problem 3

Assume that the DGP is $X \sim \text{Binomial}(n, \theta)$ with n fixed and known. Let $f(\theta) = \text{Beta}(2.5, 2.5)$. The data is as follows: $n = 100$ and $x = 39$.

(a) [easy] Find the posterior predictive distribution, $X_* \mid X$ where X_* denotes the random variable that counts the number of successes in n_* future trials.

(b) [easy] Show that for the case of predicting only $n_* = 1$ future trials, the posterior predictive distribution is $X_* \mid X \sim \text{Bernoulli}\left(\hat{\theta}^{\text{MMSE}}\right)$.

(c) [difficult] Approximate the PMF of $X_* \mid X$ as best as you can. Mark critical points and label the axes.

(d) [harder] What is the probability of $x_* \geq 10$ given your data and prior? Write your answer as a sum with terms using the beta function.

(e) [harder] [MA] Answer the previous problem exactly and then round to two decimal places using the `https://rdrr.io/snippets/` website if you don't have access to `R` on your computer. To compute $B(a, b)$ use `beta(a,b)`.

Problem 4

These are questions about Jeffreys priors.

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

(b) [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?

(c) [difficult] Explain the advantage of Jeffrey's prior in your own words.

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