MATH 341/641 Fall 2024 Homework #6

Professor Adam Kapelner

Due by email 11:59PM November 27, 2024

(this document last updated Tuesday 12th November, 2024 at 11:21pm)

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, BetaBinomial, Gamma, Poisson 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: _____

Distribution of r.v.	Quantile Function	PMF / PDF function	CDF function	Sampling Function
beta	$\texttt{qbeta}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
betabinomial	$\texttt{qbetabinom}(p, n, \alpha, \beta)$	$\texttt{d-}(x, n, \alpha, \beta)$	$\texttt{p-}(x, n, \alpha, \beta)$	$\texttt{r-}(n, \alpha, \beta)$
gamma	$\texttt{qgamma}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
ext negative binomial	$\texttt{qnbinom}(p, r, \theta)$	$\texttt{d-}(x, r, \theta)$	$\texttt{p-}(x, r, \theta)$	$\texttt{r-}(r, \theta)$
normal	$\texttt{qnorm}(p, \theta, \sigma)$	$\texttt{d-}(x, \theta, \sigma)$	$\texttt{p-}(x, \theta, \sigma)$	$\texttt{r-}(\theta, \sigma)$
inversegamma	$\texttt{qinvgamma}(p, \alpha, \beta)$	$\texttt{d-}(x, \alpha, \beta)$	$\texttt{p-}(x, \alpha, \beta)$	$\texttt{r-}(\alpha, \beta)$
Student's T	$\texttt{qt.scaled}(p, k, \mu, \sigma)$	$\texttt{d-}(x, k, \mu, \sigma)$	$\texttt{p-}(x, k, \mu, \sigma)$	$\texttt{r-}(k, \mu, \sigma)$

Table 1: Functions from $\texttt{R}$ (in alphabetical order) that can be used on this assignment and exams. The hyphen in colums 3, 4 and 5 is shorthand notation for the full text of the r.v. which can be found in column 2.

Problem 1

We will prove an interesting fact about the posterior being the result of iterative updates. Let the DGP be $\overset{iid}{\sim} \text{Bernoulli}(\theta)$ 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 (b) 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. $\mathbb{P}(\theta \mid X_2)$

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 for θ in the Bernoulli iid DGP?

(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) [harder] 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 $f(\theta \mid \boldsymbol{X})$.

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

(h) [easy] Calculuate $\hat{\theta}^{\text{MAP}}$ and $\hat{\theta}^{\text{MMSE}}$ exactly from formulas. Approximate $\hat{\theta}^{\text{MMAE}}$ 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 $\hat{\theta}^{\mathrm{MAP}}$, $\hat{\theta}^{\mathrm{MMSE}}$ and $\hat{\theta}^{\mathrm{MMAE}}$ 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).
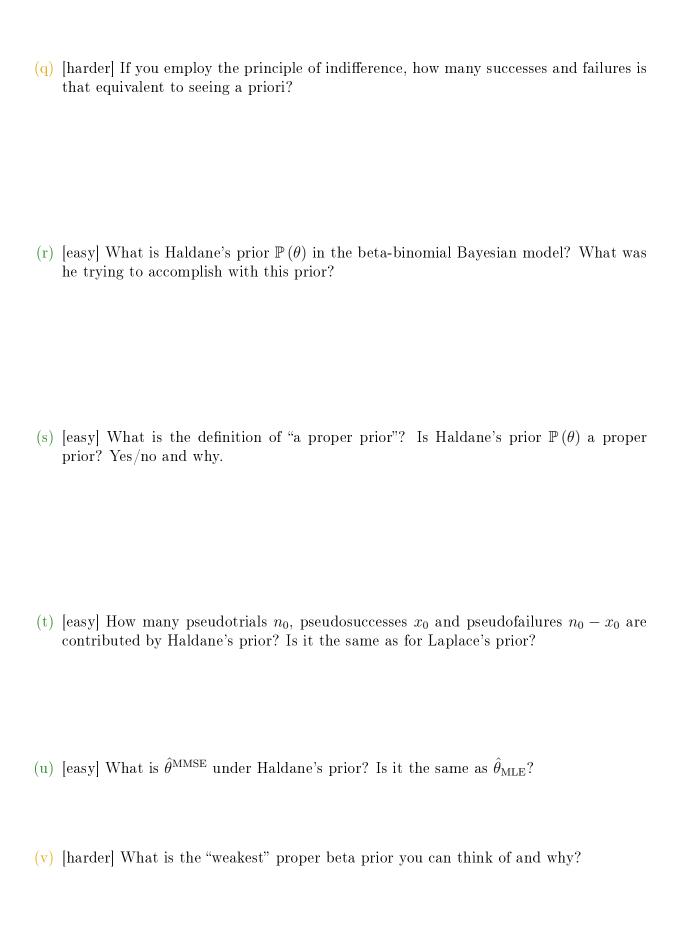
(m) [easy] Test $H_a : \theta > 0.5$ at $\alpha_0 = 5\%$. Calculate the Bayesian p-value via the `pbeta` function by using `R` on your computer (or use rdrr.io online).

(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) [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!

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

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

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

(s) [easy] What is the definition of "a proper prior"? Is Haldane's prior $\mathbb{P}(\theta)$ a proper prior? Yes/no and why.

(t) [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?

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

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

(w) [difficult] [MA] 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.

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 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 $X_1 \mid \theta,\ X_2 \mid \theta, \ldots,\ X_n \mid \theta \overset{iid}{\sim}$ Bernoulli (θ) 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 $\theta \sim$ 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

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.

9

(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] Let $n_* = 17$ for the remainder of this problem. 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 5

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] [MA] Show that the prior in the previous question is a Type IV logistic distribution and is conjugate for the binomial likelihood with the log-odds reparameterization.

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

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

Problem 6

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

(a) [harder] Prove that the Poisson likelihood for n observations, i.e. $X_1, \ldots, X_n; \theta \overset{iid}{\sim} \text{Poisson}(\theta)$, with a gamma prior yields a gamma posterior and find its parameters.

(b) [easy] Now that you see the posterior, provide a pseudodata interpretation for both hyperparameters.

(c) [harder] Find the Bayesian point estimates as function of the data and prior's hyperparameters (i.e. $\hat{\hat{\theta}}^{\mathrm{MMSE}}$, $\hat{\hat{\theta}}^{\mathrm{MMAE}}$ and $\hat{\hat{\theta}}^{\mathrm{MAP}}$).

(d) [harder] Demonstrate that $\hat{\hat{\theta}}^{\mathrm{MMSE}}$ is a shrinkage estimator and find ρ.

(e) [harder] Write the Laplace, Haldane and Jeffrey's conjugate priors below. They are all improper.

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

(g) [harder] Using the data and the prior from (f), test $H_a : \theta < 2$.

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

(i) [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.