STAT 431 — Applied Bayesian Analysis — Fall 2021

Exam

October 14, 2021

Full Name: Key ID/Email: ______________

- This is a 75 minute exam. There are 6 problems for everyone, and one additional problem for the graduate section only.
- The exam is worth a total of 50 points for the undergraduate section and 60 points for the graduate section.
- You may use your own personal notes and a standard scientific calculator. (You may *not* share these items with anyone else.) No other aids or devices are permitted!
- *Write all answers in the spaces provided.* If you require more space to write your answer, you may use the back side of the page.
- You are not allowed to communicate with anyone except the instructor or proctors before you submit this exam.

PDF	=	probability density function
PMF	=	probability mass function
MAP	=	maximum a posteriori
HPD	=	highest posterior density
PPD	=	posterior predictive distribution

1. Based on data from a small random sample, the proportion θ of U of I students who own their own computer has likelihood proportional to

$$\theta^{10}, \qquad 0 < \theta < 1$$

(a) If the sample had n students, of whom y owned their own computer, what values of n and y would produce this likelihood? (Assume a binomial data distribution.) [2 pts]

$$f(y|\theta) \propto \theta^{y}(1-\theta)^{n-y} = \theta^{10} \text{ for } 0<\theta<1$$

when $y = 10$ and $n = 10$

(b) Name a family of distributions such that any prior for θ chosen from that family would yield a posterior from that same family. [1 pt]

Beta

(c) Assume a prior for θ that is uniform on the interval $(0, 1)$.

(i) Write an expression for the kernel of the posterior density. [1 pt]

$$p(\theta|y) \propto \theta^{10}\cdot 1 = \theta^{10}, \quad 0<\theta<1$$

(ii) Precisely identify the posterior distribution: Give its name and values of the defining constants. [3 pts]

$\theta^{11-1}(1-\theta)^{1-1}$ is the kernel of Beta$(11, 1)$

(iii) Compute the *prior* mean of θ. [2 pts]

$$E(\theta) = \text{mean of Uniform}(0,1) = \frac{1}{2}$$

(iv) Compute the *posterior* mean of θ. [2 pts]

$$E(\theta|\underset{\substack{\shortparallel \\ 10}}{y}) = \text{mean of Beta}(11,1) = \frac{11}{11+1} = \frac{11}{12}$$

(d) Give an example of a prior density on $(0, 1)$ that would be improper. [2 pts]

For example,

$$\pi(\theta) \propto \frac{1}{\theta(1-\theta)} \qquad 0<\theta<1$$

(the "Haldane prior")

2. Let (L, U) be a 95% equal-tailed posterior credible interval for (scalar) parameter θ, which has posterior PDF $p(\theta \mid \boldsymbol{y})$, based on data $\boldsymbol{y}$.

(a) Give the value of each of the following: [3 pts]

$$\int_L^U p(\theta \mid \boldsymbol{y})\, d\theta = 0.95 \qquad \int_L^\infty p(\theta \mid \boldsymbol{y})\, d\theta = 0.975 \qquad \int_U^\infty p(\theta \mid \boldsymbol{y})\, d\theta = 0.025$$

(b) Which, if any, of the values in the previous part would be guaranteed to remain valid if (L, U) were instead a 95% highest posterior density (HPD) posterior credible interval? [2 pts]

Only the first one:

$$\int_L^U p(\theta \mid \boldsymbol{y})\, d\theta = 0.95$$

3. Briefly answer:

(a) Explain the difference between deterministic numerical integration and the Monte Carlo approach to approximating a posterior expected value. [4 pts]

Deterministic numerical integration is based on summing values of the integrand evaluated at fixed locations, weighted by areas (or volumes) ~~those~~ of regions those locations represent.
The Monte Carlo approach involves random simulation from the posterior, then averaging over those draws to obtain posterior expected values.

(b) Under what circumstance would the Monte Carlo approach clearly be preferred to deterministic numerical integration? Why? [2 pts]

The Monte Carlo approach scales better to high-dimensional integration regions, so it would be preferred when the parameter is relatively high-dimensional.

4. For each part below, CIRCLE the ONE BEST answer. [1 pt each]

(a) Suppose $\theta^{(1)}, \ldots, \theta^{(S)}$ is an independent sample from a posterior with PDF $p(\theta \mid y)$, where θ is scalar. Which would be the usual Monte Carlo approximation to $\mathrm{E}(\theta \mid y)$?

$\frac{1}{S}\sum_{s=1}^{S} \theta^{(s)}$ $\quad$ $\frac{1}{S}\sum_{s=1}^{S} \theta^{(s)} p(\theta^{(s)} \mid y)$ $\quad$ $\frac{1}{S}\sum_{s=1}^{S} p(\theta^{(s)} \mid y)$ $\quad$ none of these

(b) Let $Y \mid \theta \sim \text{Poisson}(\theta)$ be the data distribution. Which prior family would be conjugate for θ?

normal $\quad$ exponential $\quad$ inverse gamma $\quad$ none of these

(c) Which could be used to assess compatibility between the data and a prior distribution?

sensitivity analysis $\quad$ natural conjugacy $\quad$ posterior predictive checking $\quad$ none of these

(d) For data y and parameter θ, the expected value in the definition of Fisher information is taken with respect to which density?

$f(y \mid \theta)$ $\quad$ $p(\theta \mid y)$ $\quad$ $\pi(\theta)$ $\quad$ $m(y)$

(e) Jeffreys' prior must be

proper $\quad$ subjective $\quad$ improper $\quad$ none of these

(f) For data that are a sample from a normal distribution with known mean, the inverse gamma distribution is conjugate for the parameter representing the

precision $\quad$ standard deviation $\quad$ variance $\quad$ none of these

(g) If Y^* is a predictive quantity (unobserved potential data) based on a model for data Y with parameter θ, the usual formula for its posterior predictive density assumes which are conditionally independent?

Y and Y^* given θ $\quad$ Y^* and θ given Y $\quad$ Y and θ given Y^* $\quad$ none of these

(h) A MAP estimator is a posterior

mode $\quad$ mean $\quad$ median $\quad$ none of these

(i) Which is a type of quantile?

a median $\quad$ a mean $\quad$ a mode $\quad$ none of these

5. Briefly distinguish between Bayesian and (classical) frequentist inference, in terms of (1) how parameters are regarded, and (2) whether or not it is conditional on the data. [4 pts]

Bayesian inference regards parameters as random variables and perfoms all inference using the posterior, which is conditional on the data.

Frequentist inference regards parameters as fixed (non-random) quantities, though unknown. Since frequentist inference involves considering other possibilities for how the data may have turned out, it is not conditional on the data actually observed.

6. Consider a disease such that a randomly-chosen individual either has it ($M+$) or not ($M-$). A diagnostic test for the disease yields either a positive ($D+$) or a negative ($D-$). Let

$$a = \text{sensitivity} = \text{Prob}(D+ \mid M+) \qquad b = \text{specificity} = \text{Prob}(D- \mid M-)$$

Let π represent the proportion of individuals who actually have the disease ($0 < \pi < 1$).

(a) Express the following in terms of a, b, and π only:

(i) The probability that the randomly-chosen individual would test positive. [5 pts]

$$\begin{aligned} \text{Prob}(D+) &= \text{Prob}(D+ \mid M+)\,\text{Prob}(M+) \\ &\quad + \text{Prob}(D+ \mid M-)\,\text{Prob}(M-) \\ &= a \cdot \pi + (1-b)(1-\pi) \end{aligned}$$

(ii) The conditional probability that the randomly-chosen individual has the disease, given a positive test. [3 pts]

$$\begin{aligned} \text{Prob}(M+ \mid D+) &= \frac{\text{Prob}(D+ \mid M+)\,\text{Prob}(M+)}{\text{Prob}(D+)} \\ &= \frac{a\pi}{a\pi + (1-b)(1-\pi)} \end{aligned}$$

(b) Suppose $a + b = 1$.

(i) Show that $M+$ and $D+$ are independent (unconditionally). [3 pts]

If $a+b=1$ then $1-b=a$, so

$$\text{Prob}(M+ \mid D+) = \frac{a\pi}{a\pi + a(1-\pi)} = \frac{\pi}{\pi + (1-\pi)} = \pi$$

$$= \text{Prob}(M+)$$

$\Rightarrow$ $M+$ is indep. of $D+$

(ii) What does that imply about the usefulness of the diagnostic test? Why? [2 pts]

The test would be useless, since receiving a positive result would not change the probability of having the disease.

GRADUATE SECTION ONLY

7. A model for data value $y > 0$ has continuous scalar parameter θ, $-\infty < \theta < \infty$, and this PDF:

$$f(y \mid \theta) \;=\; \frac{1}{\sqrt{\pi}} \cdot \frac{1}{y} \cdot e^{-\theta^2} \cdot y^{2\theta} \cdot y^{-\ln y} \qquad \text{for} \quad y > 0$$

(Here, π is the mathematical constant, not a density.)

(a) Write a simplified form for the likelihood, in which all unnecessary factors are removed. [1 pt]

$$\propto e^{-\theta^2} \cdot y^{2\theta}$$

(b) Write the natural logarithm of your likelihood from the previous part. What kind of a function of θ is it? [2 pts]

$$-\theta^2 + 2(\ln y)\theta \quad + \text{constants (no } \theta)$$

which is a concave quadratic in θ

(c) Consider the flat prior $\pi(\theta) \propto 1$ for $-\infty < \theta < \infty$.

(i) Write the kernel of the posterior in the form "$\propto \exp(\cdots)$." For what values $y > 0$ is the posterior proper? [3 pts]

$$p(\theta \mid y) \propto \exp\left(-\theta^2 + (2\ln y)\theta\right) \cdot 1$$

which will be proper for all $y > 0$

(ii) Assuming a value of $y > 0$ for which it is proper, identify the posterior distribution: Name it and specify its defining constants, in terms of y. (Hint: Complete the square.) [4 pts]

$$-\theta^2 + (2\ln y)\theta = -\frac{1}{2 \cdot \frac{1}{2}}\left(\theta - \ln y\right)^2$$

$$+ \text{constants (no } \theta)$$

so

$$p(\theta \mid y) \propto \exp\left(-\frac{1}{2\left(\frac{1}{2}\right)}\left(\theta - \ln y\right)^2\right)$$

which is the kernel of Normal$\left(\ln y, \frac{1}{2}\right)$