Name: Example Solutions

STAT 535 — Intro to Bayesian Data Analysis
Test 1 — Spring 2022

1. Fill in the blanks: Bayesian posterior inference is based on a combination of prior information and data (sample) information. As the sample size increases, the data (likelihood) gets more heavily weighted in the posterior inference.

2. Suppose we have two iid observations Y_1, Y_2 that follow a gamma distribution with shape parameter known to be 3 and unknown rate parameter $\theta > 0$, i.e., having density

$$f(y|\theta) = \frac{\theta^3}{\Gamma(3)} y^{3-1} e^{-\theta y}, \quad y > 0.$$

The analyst wishes to perform inference about the unknown parameter $\theta > 0$.

(a) Write and simplify the likelihood function based on Y_1, Y_2.

$$L(\theta | y) = \prod_{i=1}^{2} \frac{\theta^3}{\Gamma(3)} y_i^2 e^{-\theta y_i}$$

$$= \left[\frac{\theta^3}{\Gamma(3)}\right]^2 \prod_{i=1}^{2} y_i^2 e^{-\theta \sum_{i=1}^{2} y_i}$$

(b) Suppose the analyst chooses a gamma prior distribution with shape parameter s and rate parameter r (given on formula sheet). Explain why this choice is sensible, with respect to the support of θ.

Since $\theta > 0$, the gamma distribution is a good choice of prior since it has support on $(0, \infty)$.

(c) If the analyst believes, before examining the data, that θ has a mean of 3 and a standard deviation of 1, then what are suitable choices for the (hyper)parameters of the gamma prior distribution? Briefly explain.

Set $\frac{s}{r} = 3$ and $\sqrt{\frac{s}{r^2}} = 1$

Solve, and see that

$\boxed{s = 9, \quad r = 3}$ satisfy this.

(d) Under this likelihood and prior, derive the posterior distribution for θ (enough to identify it, including its parameter value(s)), given Y_1, Y_2.

$$p(\theta \mid y) \propto p(\theta) L(\theta \mid y)$$

$$\propto \theta^{9-1} e^{-3\theta} \theta^{6} e^{-\theta \sum_{i=1}^{2} y_i}$$

$$= \theta^{14} e^{-\theta\left(\sum_{i=1}^{2} y_i + 3\right)}$$

which is gamma$\left(15, \sum_{i=1}^{2} y_i + 3\right)$

(e) Write an expression for the mean of the posterior distribution, and then express it explicitly as a linear combination of the prior mean and the MLE of θ, which is $\hat{\theta}_{ML} = 3/\bar{Y}$. [This one is slightly challenging (though *not* long and tedious), you may want to save it for the end.]

Posterior mean $= \dfrac{15}{\sum_{i=1}^{2} y_i + 3} = \dfrac{7.5}{\bar{Y} + 1.5}$

which can be written as

$$\underset{\text{prior mean}}{(3)}\left(\frac{1.5}{\bar{Y}+1.5}\right) + \underset{\text{MLE}}{\left(\frac{3}{\bar{Y}}\right)}\left(\frac{\bar{Y}}{\bar{Y}+1.5}\right)$$

or equivalently,

$$(3)\left(\frac{3}{\sum y_i + 3}\right) + \left(\frac{3}{\bar{Y}}\right)\left(\frac{\sum y_i}{\sum y_i + 3}\right)$$

3. Suppose we have iid observations $Y_1, \ldots, Y_n$ that follow a distribution with pdf:

$$p(y|\theta) = (y-1)\theta^2(1-\theta)^{y-2}, \quad y = 2, 3, 4, \ldots$$

where the unknown parameter is $0 < \theta < 1$. In this distribution, θ represents a success probability.

(a) Suppose you choose as a prior distribution for θ a beta(1, 3) distribution. Explain **briefly** why the beta is a reasonable choice as a prior here. Also, **briefly** explain in words what your specific choice of this beta distribution implies about your prior belief about the success probability.

The beta is a good choice because it has support on $(0,1)$. We believe the success probability is around $\frac{1}{1+3} = 0.25$.

(b) Write (and simplify as much as possible) the likelihood function $L(\theta|y_1, \ldots, y_n)$.

$$L(\theta|y) = \prod_{i=1}^{n} p(y_i|\theta) = \prod_{i=1}^{n}\left[(y_i-1)\theta^2(1-\theta)^{y_i-2}\right]$$

$$= \left[\prod_{i=1}^{n}(y_i-1)\right]\theta^{2n}(1-\theta)^{\Sigma y_i - 2n}$$

(c) Based on your prior distribution and the likelihood here, derive the posterior distribution for θ (enough to identify it specifically, including its parameter value(s)).

$$p(\theta|y) \propto p(\theta)L(\theta|y)$$

$$\propto \theta^{1-1}(1-\theta)^{3-1}\,\theta^{2n}(1-\theta)^{\Sigma y_i - 2n}$$

$$= \theta^{2n+1-1}(1-\theta)^{\Sigma y_i - 2n+3-1}, \quad \text{which is}$$

$$\boxed{\text{beta}(2n+1, \Sigma y_i - 2n + 3)}$$

(d) If we observe sample values of 9, 4, 7, 3, 5, then name the posterior distribution for θ, specifying actual numerical parameter values.

$n = 5, \quad \Sigma y_i = 28$

beta(11, 21)

(e) Based on what you know about the form of the posterior distribution here, give a Bayesian point estimate (a number) for θ using your posterior.

$$\hat{\theta}_B = \frac{11}{11+21} = \frac{11}{32} = 0.34375$$

(f) Note that $E(Y) = 2/\theta$ in this model. What is a Bayesian point estimate of $E(Y)$, given this data set?

$$\widehat{E(Y)} = \frac{2}{11/32} = \frac{64}{11} = 5.82$$

4. My wife has recently gotten into the game Wordle. Mrs. Hitchcock would like to estimate the probability that she is able to solve a randomly selected Wordle puzzle. Inspired by her devoted and, in general, awesome husband, she decides to adopt a Bayesian approach to inference. She specifies a beta prior for the unknown probability of success, and then she tries 10 puzzles (assume the outcomes are independent) and observes how many she is able to solve. A graph of the prior, (scaled) likelihood, and posterior is shown below.

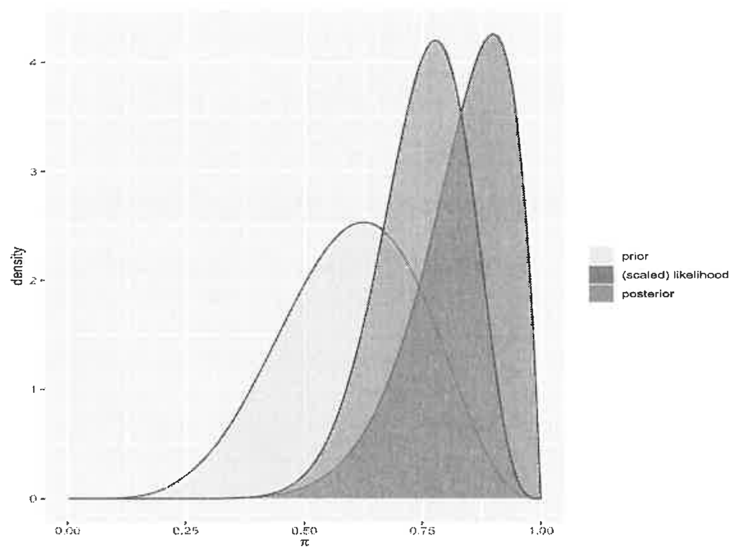

Figure 1: Prior, scaled likelihood, posterior (not necessarily in that order).

(a) Based on the plot, what is the prior that Mrs. Hitchcock has chosen? [It is one of these listed in the answer choices.]

(A) beta(2,2) (B) beta(6,4) (C) beta(4,6) (D) beta(1,9)

(b) Based on the plot, what was the observed number of successes for Mrs. Hitchcock in the 10 attempts?

(A) $y = 9$ (B) $y = 7$ (C) $y = 6$ (D) $y = 5$

(c) Using your answers to parts (a) and (b), what is a point estimate for her probability of solving a random Wordle puzzle? Indicate how you got your answer.

Posterior is beta$(6+9, 4+1) =$ beta$(15,5)$

Posterior mean is $\frac{15}{15+5} = \boxed{0.75}$

(d) In a sentence or two, discuss how her prior beliefs have been updated into the posterior information. She originally believed she had about a 60% chance of success. After seeing the data, she estimates she has a 75% chance of success.

5. Suppose for a certain company, 40% of all bolts are made in Boston and the rest are made in Chicago. We also know that 4% of bolts made in Boston are defective and 2% of bolts made in Chicago are defective.

(a) What is the probability that a randomly selected bolt is defective? Show work.

Let A = defective, B = Boston

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c)$$
$$= (.40)(.04) + (.60)(.02)$$
$$= \boxed{.028}$$

(b) If a randomly selected bolt is found to be defective, then use Bayes' Rule to find the probability that it was made in Boston. Show work.

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|B^c)P(B^c)} = \frac{(.04)(.40)}{.028}$$
$$= \boxed{0.5714}$$

6. A random sample of 24 reaction times (in seconds) of world-class sprinters from the 1996 Olympics was taken. Some R output giving the results of the analysis is given in Appendix 1 at the end of the exam. It was assumed that the reaction times were distributed $N(\mu, \sigma^2)$, and it was of interest to make inference about the mean reaction time of Olympic sprinters.

(a) What is the prior distribution for σ^2 (specifying actual numerical parameter values)?

Inv. Gamma$(66, 0.026)$

(b) What is the posterior distribution for σ^2 given $\mathbf{y}$ (specifying actual numerical parameter values)? [Requires a bit of calculator work, but the needed numbers are all in the output.]

Inv. Gamma$\left(\alpha + \frac{n}{2} - 0.5, \beta + \frac{1}{2}\left(\Sigma y_i^2 - n\bar{y}^2\right)\right)$

Inv. Gamma$\left(66 + 12 - 0.5, 0.026 + \frac{1}{2}\left(.685671 - 24(.167958^2)\right)\right)$

$\Rightarrow$ Inv. Gamma$(77.5, 0.0303)$

(c) Note that the prior for μ was:

$$\mu|\sigma^2 \sim N(\delta, \sigma^2/s_0)$$

Refer to the R code and the form of the posterior for μ. Comment on exactly how our prior beliefs about μ have been altered by observing the sample data, making reference to the values for the prior mean and posterior mean for μ.

Our prior belief was that μ was around $\delta = 0.18$, After seeing the data, our posterior estimate for μ has lowered to 0.1684 seconds.

(d) What do the two given point estimates for μ indicate about the symmetry/skewness of its posterior distribution? (Choose the best answer.)

(A) The posterior for μ is skewed, and a 95% (equal-tailed) quantile-based credible interval for μ would be the same as a 95% HPD credible interval for μ.

(B) The posterior for μ is symmetric, and a 95% (equal-tailed) quantile-based credible interval for μ would be the same as a 95% HPD credible interval for μ.

(C) The posterior for μ is skewed, and a 95% (equal-tailed) quantile-based credible interval for μ would be different than a 95% HPD credible interval for μ.

note the posterior mean and median are the same $\Rightarrow$ symmetric posterior.

(D) The posterior for μ is symmetric, and a 95% (equal-tailed) quantile-based credible interval for μ would be different than a 95% HPD credible interval for μ.

(e) In a sentence, briefly but carefully interpret (in the context of the variable in the study) the given 95% credible interval for μ.

With posterior probability 0.95, the mean reaction time for Olympic sprinters is between 0.1606 and 0.1763 seconds.

Appendix 1

```
> react <- c(.187,.152,.137,.175,.172,.165,.184,.185,.147,.189,.172,.156,.168,.140,.214,.163,.202,.173,.175,.154,.160,.169,.148,
> y<- react
> ybar <- mean(y); n <- length(y); my.s <- sd(y)
>
> # prior parameters
> my.alpha <- 66; my.beta <- 0.026
>
> # prior parameters
> my.delta <- 0.18; s0 <- 1
>
> ybar
[1] 0.1679583
> n
[1] 24
> sum(y^2)
[1] 0.685671
> my.s
[1] 0.0193716
>
> library(pscl)  # loading pscl package
> library(TeachingDemos)  # loading TeachingDemos package
>
> ### Point estimates for sigma^2:
>
> p.mean.sig.sq <- (my.beta + 0.5*(sum(y^2) - n*(ybar^2)) ) / (my.alpha + n/2 - 0.5 - 1)
> p.median.sig.sq <- qigamma(0.50, my.alpha + n/2 - 0.5, my.beta + 0.5*( sum(y^2) - n*(ybar^2) ) )
>
> print(paste("posterior.mean for sigma^2=", round(p.mean.sig.sq,4),
+       "posterior.median for sigma^2=", round(p.median.sig.sq,4) ))
[1] "posterior.mean for sigma^2= 0.0004 posterior.median for sigma^2= 0.0004"
>
> ### Marginal Interval estimate for sigma^2:
>
> hpd.95.sig.sq <- hpd(qigamma, alpha=my.alpha + n/2 - 0.5, beta=my.beta + 0.5*( sum(y^2) - n*(ybar^2) ) )
> hpd.95.sig.sq
[1] 0.0003111014 0.0004875030
>
> #### AN APPROACH FOR INFERENCE ABOUT mu:
>
> # Randomly sample many values for the posterior of sigma^2:
>
> sig.sq.values <- rigamma(n=1000000,alpha=my.alpha + n/2 - 0.5, beta=my.beta + 0.5*( sum(y^2) - n*(ybar^2) ) )
>
> # Randomly sample many values from the posterior of mu, GIVEN the sampled values of sigma^2 above:
>
> mu.values <- rnorm(n=1000000,mean=((sum(y)+my.delta*s0)/(n+s0)), sd=sqrt(sig.sq.values/(n+s0)) )
>
> # Point estimates for mu:
>
> print(paste("posterior.mean for mu=", round(mean(mu.values),4),
+       "posterior.median for mu=", round(median(mu.values),4) ))
[1] "posterior.mean for mu= 0.1684 posterior.median for mu= 0.1684"
>
> # 95% HPD interval estimate for mu:
>
> round(emp.hpd(mu.values),4)
[1] 0.1606 0.1763
>
```