

Theses answers come with no warranty.

Birds

A. [5%]

To get the full score, one must simply write:

$$p(\theta = S|y) = \frac{f(y|\theta = S)\pi(\theta = i)}{\sum_{i \in \{S, F, B\}} f(y|\theta = i)\pi(\theta = i)}$$

with the exact specification of $\theta = S$ and the summation in the denominator.

B. [5%]

The most reasonable prior is given by expert 3 because the elicited prior is supported by a lot of previously obtained data. Option 1 could be accepted if the student feels that option 3 does not guarantee enough confidence. Option 2 is not acceptable because, as seen in class, it is not good practise to use string elicited priors that limit the parameter space.

Assuming option 3, then the prior is defined as:

$$\pi(\theta = S) = 0.50$$

$$\pi(\theta = F) = 0.30$$

$$\pi(\theta = B) = 0.20$$

Assuming option 1, then the prior is defined as:

$$\pi(\theta = S) = \frac{1}{3}$$

$$\pi(\theta = F) = \frac{1}{3}$$

$$\pi(\theta = B) = \frac{1}{3}$$

C. [20%]

Assuming option 3, then the two models have the same prior probability. Therefore,

$$BF = \frac{p(M_1|y)}{p(M_2|y)}$$

with

$$\begin{aligned} p(M_1|y) &= \frac{((0.7)^2 * 0.5)}{((0.7)^2 * 0.5) + ((0.2)^2 * 0.3) + ((0.1)^2 * 0.2)} = \\ &= \frac{(0.49 * 0.5)}{(0.49 * 0.5) + ((0.04 * 0.3) + (0.01 * 0.2))} = \\ &= \frac{0.245}{0.245 + 0.012 + 0.002} = \frac{0.245}{0.259} \approx 0.95 \end{aligned}$$

Likewise

$$p(M_2|y) = \frac{((0.04 * 0.3) + (0.01 * 0.2))}{0.259} = \frac{0.012 + 0.002}{0.259} = \frac{0.0122}{0.259} \approx 0.05$$

Then

$$BF = \frac{0.95}{0.05} \approx 20$$

A BF of 20 is indicative of positive-to-strong support for model M_1 .

D. [5%]

To get the full score, one must simply write:

$$p(\theta = S|y) = \frac{\sum_{j \in \{-1, 0, 1\}} f(y|\theta = S)\pi(\theta = i|\tau = j)}{\sum_{j \in \{-1, 0, 1\}} \sum_{i \in \{S, F, B\}} f(y|\theta = i)\pi(\theta = i|\tau = j)}$$

with the exact specification of all summations.

Extinctions

(i)

Here we need to use Bayes formula. By dropping all terms that do not contain λ , we realise that the posterior is a Gamma distribution and, after some algebra, we find that $\alpha' = \alpha + 1$.

$$\begin{aligned} p(\lambda|x) &\approx p(x|\lambda)p(\lambda|\alpha, \beta) \\ p(\lambda|x) &\approx \lambda e^{-\lambda x} \lambda^{\alpha-1} e^{-\lambda/\beta} \\ p(\lambda|x) &\approx \lambda^{\alpha-1+1} e^{\lambda x - \lambda/\beta} \end{aligned}$$

for the latter we evince $p(\lambda|x) = G(\alpha', \beta')$ with $\alpha' = \alpha + 1$.

(ii)

Given the expected value and variance of a Gamma distribution, a suitable prior might be $G(\alpha = 1, \beta = 3.5)$, which produces the desired expected value with a large variance.

With this choice the posterior mean is the expected value of $G(1+1, 3.5+2.5)$ which is 12.

(iii)

This bayes factor provides very strong support for M1, and it represents the shift in odds when we move from the prior towards the mean.

Given the HPD, we can say that the posterior is lower than 0.05.

Salamanders

(i)

Although not required to answer this question, students may provide the formula for the posterior distribution:

$$P(\theta|y) = N(\theta | \frac{\sigma^2 \mu + \tau^2 y}{\sigma^2 + \tau^2}, \frac{\sigma^2 \tau^2}{\sigma^2 + \tau^2})$$

Even without knowing this formula, students should know that if the variance of the likelihood is much larger than the variance of the prior, then the expected value of the posterior will approach the mean of the prior.

Likewise, students should be able to recall that the variance of the posterior here is smaller than the variance of either the likelihood and the prior.

Here I am testing what learned on the Normal/Normal model and the interplay between data and prior. We discussed it extensively in class.

(ii)

Given these observations, we can assume that a reasonable prior should have μ close to 5.1 with a variance that allows the distribution to span the whole range. Students should not focus their choices on the exact range and extreme percentiles. Rather, given that our data will be large and therefore dominating the posterior, it is reasonable to choose a relatively large variance to make sure to cover beyond the observed range. For instance a variance of 4 should work as we know that the mean plus/minus three times the standard error roughly covers the Normal distribution. Students may also wish to specify the prior should be defined only for non-zero positive values.

Here I am testing how students can translate some prior information into a suitable probability distribution based on some good practice we discussed in class.

(iii)

A Bayes factor of 10 suggests minimal if no support for $P(\theta|y) \geq 7$. This may also mean that there has been a small shift from the prior to the posterior towards M1. With such HDPI, the probability of being lower than 1.5 is 0.10.

Here I test the understanding of the interpretation of Bayes factor on the posterior, not on the prior itself. I also test whether students understood the interpretation of HDPI as proper probabilities (unlike frequentist approach) and the fact that a Normal distribution is symmetric.

(iv)

Students should be able to recognise that λ and θ are independent when conditional on μ . Therefore, $P(y, \theta, \mu, \lambda, \rho) = P(\rho)P(\mu)P(\theta|\rho, \mu)P(\lambda|\mu)P(y|\theta)$.

Here I test the ability to recognise conditional independencies and use the chain rule. We discussed a similar example in class.