ST440/540 – 2018 Mid-term 1 – Solutions

1. In the early twentieth century, it was generally agreed that Hamilton and Madison (ignore Jay for now) wrote 51 and 14 of the Federalist Papers, respectively. There was dispute over how to attribute 12 other papers between these two authors. In the papers attributed to Madison, the word "upon" was used 0.23 times per 1,000 words, compared to 3.24 per 1,000 words for Hamilton.

   (a) Give a reasonable prior probability that a disputed paper was written by Hamilton.

   ANS: The prior probability that a paper was written by Hamilton is $p = 51/(51 + 14)$.

   (b) If the word "upon" is used three times in a disputed text of length 1,000 words, what is the posterior probability it was written by Hamilton (your answer can be a formula)?

   ANS: Let $\theta_M = 0.23/1000$, $\theta_H = 3.24/1000$, $n = 1,000$ and $Y = 3$. The unknown quantity is $A = \{M, H\}$, where "A" is for author, "M" is for Madison and "H" is for Hamilton. The data can be modeled as $Y|A \sim \text{Binomial}(n, \theta_A)$ and the prior is $\text{Prob}(A = H) = p$. Bayes rule then gives

$$\begin{aligned}
\text{Prob}(A = H|Y) &= \frac{f(Y|H)p}{f(Y|H)p + f(Y|M)(1-p)} \\
&= \frac{\theta_H^Y(1-\theta_H)^{n-Y}p}{\theta_H^Y(1-\theta_H)^{n-Y}p + \theta_M^Y(1-\theta_M)^{n-Y}(1-p)} = 0.998.
\end{aligned}$$

   (c) Give one assumption you are making in (b) that is likely unreasonable. Justify your answer.

   ANS: By assuming a binomial distribution, we are assuming the words are independent, which is unreasonable because, e.g., it is impossible to see two consecutive "upon"'s.

   (d) In (b), if we changed the number of "upon"'s to one, do you expect the posterior probability to increase, decrease or stay the same? Why?

   ANS: The "upon" count is now closer to Madison's expected usage, and so the posterior probability should decrease.

   (e) In (b), if we changed the text length to $10,000$ words the number of "upon"'s to 30, do you expect the posterior probability to increase, decrease or stay the same? Why?

   ANS: The sample proportion is the same (still closer to Hamiltons rate) and so with more data we can be more definitive, and so the posterior probability should increase.

   (f) In (b), if we allowed for the possibility of a third author (Jay), do you expect the posterior probability to increase, decrease or stay the same? Why?

   ANS: The prior probability would decrease and the likelihood would stay the same, so the posterior probability should decrease.

2. Say Y follows the Laplace distribution with PDF

$$f(y|\tau) = \frac{\tau}{2}\exp(-\tau|y|).$$

Identify a conjugate prior for the variance parameter $\tau > 0$ and derive the resulting posterior (assuming a $n = 1$ observation).

ANS:If $\tau \sim \text{Gamma}(a, b)$ then

$$\begin{aligned} f(\tau|Y) &\propto f(Y|\tau)f(\tau) \\ &\propto [\tau\exp(-\tau|Y|)]\left[\tau^{a-1}\exp(-\tau b)\right] \\ &\propto \tau^{(1+a)-1}\exp[-(|Y|+b)\tau]. \end{aligned}$$

Therefore the gamma family is conjugate and the posterior is $\tau|Y \sim \text{Gamma}(a+1, |Y|+b)$.

3. The web developer in your company has a new idea to tweak your company's online ad. She claims if you add "Dilly Dilly" you will get more clicks. Your CEO decides to green light an experiment. You collect data for 14 days with and without the phrase, and the record the number of clicks received each day. Describe how you would conduct a Bayesian analysis of the data from this experiment. In particular, give your prior, posterior, and rule for determining if the new ad is preferred to the current ad.

ANS: Let $N = 14$, Y_1 be the total number of clicks in the 14 days with "Dilly Dilly", and Y_2 be the total number of clicks on the other days. My model is

$$Y_j \sim \text{Poisson}(N\lambda_j) \quad \text{where} \quad \lambda_j \sim \text{Gamma}(\epsilon, \epsilon)$$

independent over j. To give an uninformative prior I would select $\epsilon = 0.01$. Then the posterior (independent over j is)

$$\lambda_i|Y_j \sim \text{Gamma}(Y_j + \epsilon, N + \epsilon).$$

I would conclude that the new ad is more effective than the old one if

$$\text{Prob}(\lambda_1 > \lambda_2|Y_1, Y_2) > 0.9,$$

which I would approximate with Monte Carlo simulation.

4. Say there is $n = 1$ observation with $Y|\mu \sim \text{Normal}(\mu, 1)$ and the prior is $\mu \sim \text{Normal}(0, 1)$. The objective is to test the null hypothesis that $H_0 : \mu \leq 0$ versus the alternative hypothesis that $H_1 : \mu > 0$.

   (a) Give the posterior of μ.

   ANS: The formula from class with $n = m = \sigma = 1$ and $\theta = 0$ gives $\mu|Y \sim \text{Normal}(Y/2, 1/2)$.

   (b) Give an expression for the posterior probability of the null hypothesis. Your answer should be a function of the standard normal CDF, Φ, and/or, the standard normal quantile function, Q (i.e., the function so that $Q[\Phi(x)] = x$).

   ANS: $\text{Prob}(H_0|Y) = \text{Prob}(\mu < 0|Y) = \text{Prob}(Z < \frac{-Y/2}{\sqrt{1/2}}|Y) = \text{Prob}(Z < -Y/\sqrt{2}|Y) = \Phi(-Y/\sqrt{2})$.

   (c) Give the (frequentist) definition of the Type I error rate.

   ANS: The Type I error rate is long-run probability of rejecting the null hypothesis assuming it is true.

   (d) Say that our rule is to reject the null hypothesis if its posterior probability is less than τ. Explain how you would select τ so that the test has Type I error rate equal to 0.05. Give the answer in terms of Φ and Q.

   ANS: We reject the null if $\Phi(-Y/\sqrt{2}) < \tau$, i.e., if $Y > -\sqrt{2}Q(\tau)$. Assume the null hypothesis is true and thus $\mu = 0$ and $Y \sim \text{Normal}(0, 1)$. Then the Type I error rate is $\text{Prob}[Y > -\sqrt{2}Q(\tau)] = \text{Prob}[Y < \sqrt{2}Q(\tau)] = \Phi[\sqrt{2}Q(\tau)]$. Setting this to 0.05 and solving for τ gives $\tau = \Phi[Q(0.05)/\sqrt{(2)}] \approx 0.12$.