Quiz 4

Dustin Leatherman

October 14, 2020

1 1

The table below gives the number of Federalist papers written by Hamilton, Madison, and Jay and the number of times they used the word "upon" in per 1,000 words in their papers. Say there is a newly uncovered paper with Y = 7 instances of "upon" in n = 10,000 words.

Author	Number of Papers	"Upon" per 1000 words
Hamilton	51	3.24
Madison	14	0.23
Jay	5	5.61
Total	70	2.81

1.1 a

Using the proportion of the 70 pages assigned to each author as the prior (so P(Hamilton) = 51/70), compute the posterior probability that the newly uncovered paper was written by Hamilton

Author	Number of Papers	"Upon" per 1000 words	P(Y given θ)
Hamilton	51	3.24	0.00324
Madison	14	0.23	0.00023
Jay	5	5.61	0.00561
Total	70	2.81	0.00281

$$
\begin{aligned}
P(H|Y=7) =& \frac{P(Y=7|H)P(H)}{P(Y=7|H)P(H)+P(Y=7|J)P(J)+P(Y=7|M)P(M)} \\
=& \frac{0.00324 \cdot (51/70)}{0.00324 \cdot (51/70) + 0.00023 \cdot (14/70) + 0.00561 \cdot (5/70)} \\
=& 0.8033
\end{aligned} \tag{1}
$$

1.2 b

> We have used the word "upon" to assign papers to authors, but other words could be used ("The", "impeachment", "liberty", etc.). If you could only pick one word to use, how would you pick the word?

I would compute the Term Frequency Inverse-Document Frequency (TF-IDF) score for a series of papers for the relevant authors and choose the highest term that is common across the relevant authors. This would ensure that participles and other common terms such as "I", "The", "upon", etc. are not considered. Using common terms makes it more difficult to pin an article or style to a particular individual since any author could use "upon".

2 2

> Assume $Y|\theta \sim Gamma(1, \theta)$ and prior distribution $\theta \sim Gamma(a, b)$

2.1 a

> Derive the posterior distribution of θ

$$
\begin{aligned}
f(\theta|Y_1,...,Y_n) =& \frac{f(Y_1,...,Y_n|\theta)f(\theta)}{f(Y_1,...,Y_n)} \\
=& \frac{f(Y_1,|\theta)\cdot f(Y_2,|\theta)\cdot ... \cdot f(Y_n,|\theta)\cdot f(\theta)}{f(Y_1,...,Y_n)} \\
\propto& \theta^n e^{-\theta\sum_{i=1}^n y_i}\cdot \frac{b^a}{\Gamma(a)}\theta^{a-1}e^{-b\theta} \\
\propto& \theta^{a+n-1}e^{-\theta(b+\sum_{i=1}^n y_i)} \\
\theta|Y_1,...,Y_n \sim& Gamma(a+n, b+\sum_{i=1}^n y_i)
\end{aligned} \tag{2}
$$

2.2 b

Argue that the gamma prior is conjugate.

The result produced in (2) is a PDF that has an appearance similar to a Gamma Distribution and as such, the Gamma distribution can be derived from the equation.