JSC270 - Assignment 3

Cindy (Shih-Ting) Huang

March 2021

Assignment 3 Github Repo: https://github.com/cindyellow/JSC270_Assignment3

Question 1

A) We are given that the likelihood function of n individuals is

$$f(x_1, \ldots, x_n|\theta) = \theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i}$$

and the prior distribution is

$$f(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}$$

In order to determine $f(\theta|x)$, we utilize the Bayes' Rule which explains that:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_\theta f(x|\theta')f(\theta')\,d\theta'}$$

We first determine the numerator:

$$\begin{aligned} f(x|\theta)f(\theta) &= [\theta^{\sum\limits_{i=1}^{n} x_i}(1-\theta)^{n-\sum\limits_{i=1}^{n} x_i}][\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a-1}(1-\theta)^{b-1}] \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a+(\sum\limits_{i=1}^{n} x_i)-1}(1-\theta)^{b+n-(\sum\limits_{i=1}^{n} x_i)-1} \end{aligned}$$

Next, we determine the normalizing constant in the denominator:

$$\begin{aligned} \int_\theta f(x|\theta')f(\theta')\,d\theta' &= \int_0^1 f(x|\theta')f(\theta')\,d\theta' && \text{(since } 0 < \theta' < 1\text{)} \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\int_0^1 \theta'^{a+(\sum\limits_{i=1}^{n} x_i)-1}(1-\theta')^{b+n-(\sum\limits_{i=1}^{n} x_i)-1}\,d\theta' \\ &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}Beta(a+\sum_{i=1}^{n} x_i, b+n-\sum_{i=1}^{n} x_i) \\ & \qquad \text{(Since } Beta\text{(a,b)'s distribution is } \textstyle\int_0^1 \theta^{a-1}(1-\theta)^{b-1}\,d\theta) \end{aligned}$$

Combining the two results, we get:

$$f(\theta|x) = \frac{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\theta^{a+(\sum_{i=1}^{n} x_i)-1}(1-\theta)^{b+n-(\sum_{i=1}^{n} x_i)-1}}{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}Beta(a+\sum_{i=1}^{n} x_i, b+n-\sum_{i=1}^{n} x_i)}$$

$$= \frac{\theta^{a+(\sum_{i=1}^{n} x_i)-1}(1-\theta)^{b+n-(\sum_{i=1}^{n} x_i)-1}}{\frac{\Gamma(a+\sum_{i=1}^{n} x_i)\Gamma(b+n-\sum_{i=1}^{n} x_i)}{\Gamma(a+b+n)}} \qquad \text{(Since } Beta(\text{a,b}) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}\text{)}$$

$$= \frac{\Gamma(a+b+n)}{\Gamma(a+\sum_{i=1}^{n} x_i)\Gamma(b+n-\sum_{i=1}^{n} x_i)}\theta^{a+(\sum_{i=1}^{n} x_i)-1}(1-\theta)^{b+n-(\sum_{i=1}^{n} x_i)-1}$$

which is the distribution for $Beta(a+\sum_{i=1}^{n} x_i, b+n-\sum_{i=1}^{n} x_i)$.

B)

$$E(\theta|x) = \frac{a+\sum_{i=1}^{n} x_i}{a+(\sum_{i=1}^{n} x_i)+b+n-(\sum_{i=1}^{n} x_i)}$$

(Since our posterior distribution is $Beta(a+\sum_{i=1}^{n} x_i, b+n-\sum_{i=1}^{n} x_i)$)

$$= \frac{a+\sum_{i=1}^{n} x_i}{a+b+n}$$

C) $Beta(1{,}1) = \theta^0(1-\theta)^0 = 1 \sim Uniform(0,1)$ because the distribution for that is f(x) $= \frac{1}{1-0} = 1$. Using this special prior distribution, we are suggesting that all possible values of θ are equally likely.

D) The likelihood estimate ($\hat{\mu} = 0.2$) from our sample dataset is lower than the posterior mean ($\mu \approx 0.227$), which is lower than the prior mean($\mu \approx 0.500$). This is because the posterior distribution represents our updated model based on the prior distribution accounted for the data we observed. Since the mean of the sample data is lower than the prior mean, we know that the posterior mean will reflect the pull of the prior mean towards the likelihood estimate, effectively evaluating to a value between the two.

E) Both the prior ($\mu \approx 0.249$) and posterior means ($\mu \approx 0.209$) have lowered, with the former observing a more significant decrease. If we look at prior and posterior means as calculated by their respective expectation, we find that in both distributions, the expectation is indirectly proportional to b, meaning that a greater value of b as in 1e) will cause the means to be lower.

F) The prior mean is approximately the same as in 1e), but the posterior mean has increased to $\mu \approx 0.312$, which is higher than in both 1d) and 1e). This occurs because although we have the same a and b values as 1e), the observed COVID infection has increased from 4 to 12, effectively increasing the likelihood estimate to $\hat{\mu} = 0.6$. Hence, the prior mean remains the same while the posterior mean will increase as our prior mean gets accounted for the higher estimate in its calculation.

Question 2

(A)

$$
\begin{aligned}
E(X) &= \int_{-\infty}^{\infty} xf(x)\,dx && (1)\\
&= \int_{0}^{\infty} x\lambda e^{-\lambda x}\,dx && (2)\\
&= \lambda\int_{0}^{\infty} xe^{-\lambda x}\,dx && (3)\\
&= \lambda(\frac{-1}{\lambda}xe^{-\lambda x}]_0^\infty + \int \frac{1}{\lambda}e^{-\lambda x}\,dx) && (4)\\
&= \lambda(\frac{-1}{\lambda}xe^{-\lambda x} + (\frac{1}{\lambda})(\frac{-1}{\lambda})e^{-\lambda x})]_0^\infty && (5)\\
&= -xe^{-\lambda x} - (\frac{1}{\lambda})e^{-\lambda x}]_0^\infty && (6)\\
&= \frac{x}{-e^{\lambda x}}]_0^\infty - \frac{1}{\lambda e^{\lambda x}}]_0^\infty && (7)\\
&= 0 - (0 - \frac{1}{\lambda}) = \frac{1}{\lambda} && (8)
\end{aligned}
$$

Hence, the expectation of X when $\lambda = 4$ is $\frac{1}{4}$.

(B) See code.

(C) As the sample size grows, the sample mean approaches the theoretical mean $\mu = 0.5$. This behavior corresponds to the phenomenon described by the Law of Large Number.

Question 3

(A) See code.

(B) See code.

(C) See code.

(D) See code.

(E) The performance of my logistic classifier from 3A) can be described by its test set accuracy 0.985. In comparison with that, the out-of-bag accuracy from the bootstrap method is lower at 0.972. I think that the logistic classifier performs marginally better because for our calculation of y_i, we start off with a regression line equation, more or less indicating a line-like relationship between the input and output. Decision trees are not as suitable for this purpose since usually there is a less linear association. Additionally, decision trees would perform well when there are multiple features we can split on. In this case, we are only feeding one feature (x_i) into the model, so the number of splits it can do is relatively more limited.

Question 4

i) The average price of a cup of coffee in Toronto can be best represented by Graph (1) displaying a Gamma distribution. Firstly, we see that the distribution is skewed towards one end. Given the scenario, this is valid because one cup of coffee is usually more expensive than if you buy multiple, for which one often gets more discount.

ii) The proportion of individuals who catch a cold in February can be modeled by Graph (3)'s Beta distribution. There are several characteristics that can explain this. We observe that the domain of x is between 0 and 1 as should proportion of a population be. Also, the curve is negatively sloped since the probability that a higher proportion of people has caught a cold is lower, eventually approaching $f(x) = 0.0$ for $x = 1.0$, the all individuals have caught a cold, which is nearly impossible.

iii) The number of people who own a dog can be modeled by Graph (4). This is because a Binomial distribution stems from experiments of Bernoulli trials, which only have binary results. This makes sense because each person in the small group either owns or doesn't own a dog.

iv) The average weekly change in US unemployment can be represented by Graph (2)'s Normal distribution. First, we note that x ranges between negative and positive values, which is reasonable since unemployment can experience both negative and positive change. Moreover, the scenario corresponds to a bell-shaped curve because when measured on a large scale, most observations are close to the mean rather than in the extremes on either side.