ISYE 6420 Bayesian Statistics

Midterm

October 26, 2021

1.

We have the following information provided from the problem,

A: Probability that Chad get's an A

C: Probability that Chad get's a car

CB: Probability that Chad goes to Coco Beach

$P(A) = 0.7$

$P(A^C) = 0.3$

$P(C|A) = 0.8$

$P(C^C|A) = 0.2$

$P(C|A^C)$

$= 0.1$

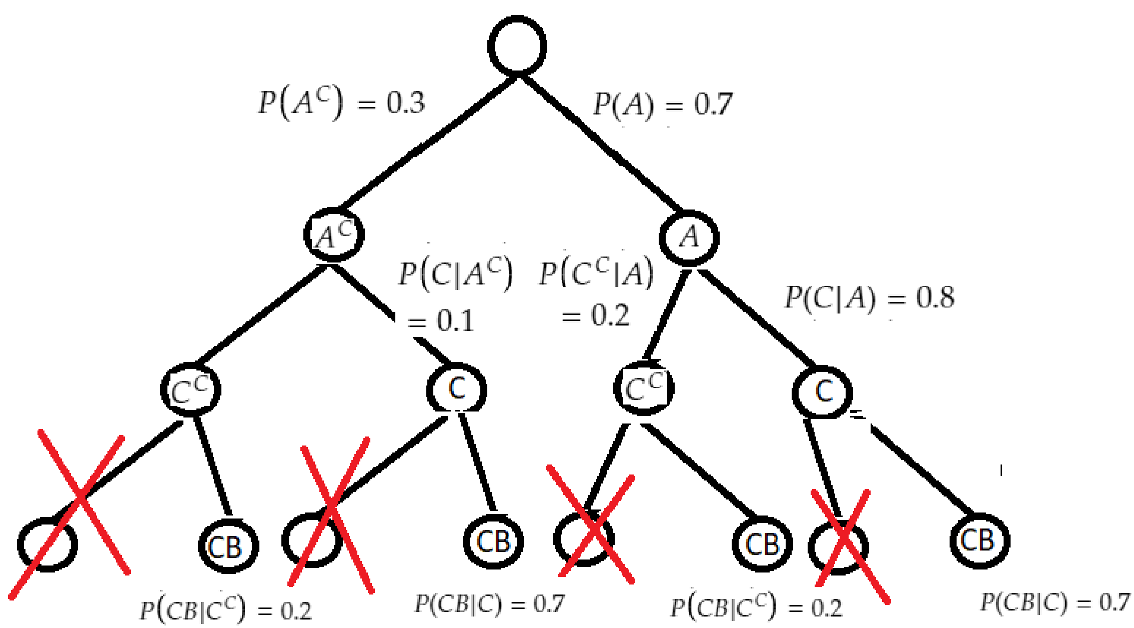
$P(C^C|A^C)$

$= 0.9$

$P(CB|C) = 0.7$

$P(CB|C^C) = 0.2$

When taking these probabilities and forming a tree, we can visualize the probablities required to get our prior for CB P(CB):



$$
\begin{aligned}
P(CB) = \; & P(CB|C) * P(C|A) * P(A) \; + \\
& P\left(CB|C^C\right) * P\left(C^C|A\right) * P(A) \; + \\
& P(CB|C) * P\left(C|A^C\right) * P\left(A^C\right) \; + \\
& P(CB|C) * P\left(C^C|A^C\right) * P\left(A^C\right)
\end{aligned}
$$

$$
\begin{aligned}
= \; & 0.7 * 0.8 * 0.7 \; + \\
& 0.2 * 0.2 * 0.7 \; + \\
& 0.7 * 0.1 * 0.3 \; + \\
& 0.7 * 0.9 * 0.3
\end{aligned}
$$

$$
\begin{aligned}
&= 0.392 \; + \; 0.028 + \; 0.021 + \; 0.189 \\
&= 0.63
\end{aligned}
$$

$$
\begin{aligned}
P(C) &= P(C|A) * P(A) \; + P\left(C|A^C\right) * P\left(A^C\right) \\
&= 0.8 * 0.7 \; + \; 0.1 * 0.3 \\
&= 0.59
\end{aligned}
$$

$$
\begin{aligned}
P(C|CB) &= \frac{P(CB|C) * P(C)}{P(CB)} \\
&= \frac{0.7 * 0.59}{0.63}
\end{aligned}
$$

$$= 0.65 \textit{ repeated or } 65\% \textit{ chance he has a car given that he went to Coco Beach}$$

2.

Based on our likelihood being normal we have the following:

$y_i|\theta \sim iid\ N\left(\theta, \sigma^2\right)$

$\theta \sim U(0, 1)$

$\frac{1}{2}\, exp(-\frac{1}{8}\left((y_i - \Theta')^2 - (y_i - \Theta)^2\right)$

$$f(y_i|\theta) \; = \; \frac{1}{\left(2\pi\sigma^2\right)^{\frac{\pi}{2}}}\, exp\left(\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2\right)$$

Taking the logs of each side we can get the following:

$$L(\theta) \ = \ -\frac{n}{2}\, log_e[2\pi] - \frac{n}{2}\, log_e[\sigma] - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2$$

Since we only care about terms with respect to θ, the first two terms can simplify to a constant c:

$$L(\theta) \ = \ c - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \theta)^2$$

When differentiating twice we get the following:

$$\frac{dL(\theta)}{d\theta} \ = \ \frac{1}{\sigma^2}\left(\sum_{i=1}^{n}(y_i - \theta n)\right)$$

$$\frac{d^2L(\theta)}{d\theta^2} \ = \ \frac{-n^2}{\sigma^2}\frac{(\theta_0 - \theta)^2}{2}$$

When we take the Taylor's expansion of the above equation gives me the following:

$$= \frac{-n^2}{\sigma^2}\frac{(\theta_0 - \theta)^2}{2}$$

We can then rearrange this equation and plug back in the exponent terms, and we can see that the posterior distribution for θ follows a normal distribution:

$$f(\theta) \ = k \ exp\ \frac{(y_i - \theta)^2}{2\left(\frac{\sigma}{\sqrt{n}}\right)^2}$$

3.

We have 197 animals are distributed into four categories:

$$\begin{aligned} y \ &= \ (y_1,\ y_2,\ y_3,\ y_4) \\ &= \ (125,\ 18,\ 20,\ 34) \end{aligned}$$

When taking into account our 4 categories and the cell probabilities, we can observe that the likelihood is the following:

Probability mass function [edit]

In general, if the random variable X follows the binomial distribution with parameters $n \in \mathbb{N}$ and $p \in [0,1]$, we write $X \sim B(n, p)$. The probability of getting exactly k successes in n independent Bernoulli trials is given by the probability mass function:

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$p(y|\theta) \ \propto \ (2+\theta)^{y1}(1-\theta)^{y2+y3}\theta^{y4}$$

$Since\ the\ prior\ for\ \theta\ is\ =\ 1\ ,\ we\ know\ from\ this\ that\ our\ posterior\ is:$

$$p(\theta|y) \propto \ p(y|\theta)p(\theta)$$
$$p(\theta|y) \ \propto \ (2+\theta)^{y1}(1-\theta)^{y2+y3}\theta^{y4}$$

$After\ we\ split\ the\ first\ cell\ into\ 2,\ we\ know\ have\ the\ following\ 5\ categories:$

$$(y,\ y_0)\ \ =\ (125\ -\ y_0, y_0,\ 18,\ 20,\ 34)$$

$Followed\ by\ the\ probabilties\ that\ are\ given:$

$$\left(\frac{1}{2}, \frac{\theta}{4}, \frac{1-\theta}{4}, \frac{1-\theta}{4}, \frac{\theta}{4}\right)$$

$From\ this\ we\ can\ derive\ that\ p(y_0, y|\theta)\ \propto (1-\theta)^{38}\theta^{y_0+34}$

$Where\ we\ can\ see\ that\ this\ is\ starting\ to\ resemble\ the\ Beta\ distribution,$

$where:\ Beta(y_0+35,\ 39)$
$We\ can\ use\ this\ to\ get\ derive\ the\ joint\ distribution\ betwee\ y_0\ and\ \theta:$

$$p(y_0, \theta|y)\ \propto\ \frac{197!}{(125-y_0)!y_0!}\left(\frac{\frac{1}{2}}{\frac{1}{2}+\frac{\theta}{4}}\right)^{125-y_0}\left(\frac{\frac{\theta}{4}}{\frac{1}{2}+\frac{\theta}{4}}\right)^{y_0}$$

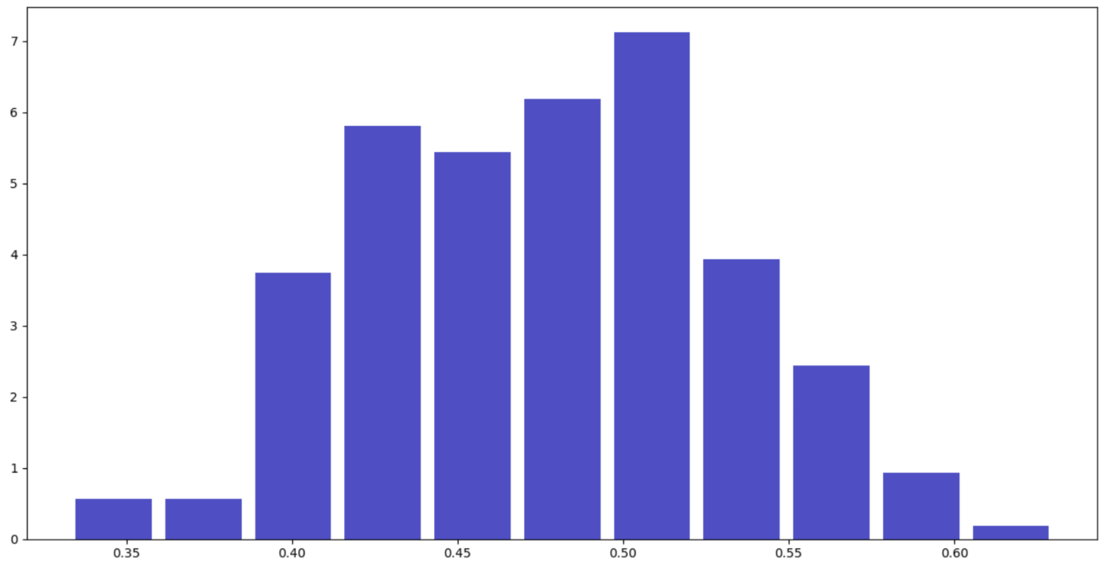
$We\ can\ see\ that\ this\ equation\ is\ starting\ to\ resemble\ the\ pmf\ of\ a\ binomial\ distribution:$

$$f(k,n,p) = \Pr(k;n,p) = \Pr(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Where: $Binom\left(125, \frac{\theta}{2+\theta}\right)$

```
import numpy as np
from scipy.special import gamma
from scipy.stats import norm
import matplotlib.pyplot as plt

np.random.RandomState(1)

#Set the number of iterations
n = 197
#Set the empty lists to hold the updated values for theta and tau
for each variable (i.e 1,2)
thetas1 = []
y = []
a = 35
b = 39

# start, initial values
y0 = 0
theta1 = 0

for i in np.arange(1,n+1).reshape(-1):
    #Determine the value of updated value
    updatedTheta1 = np.random.beta(theta1+a,b, size=1)[0]
    p  = updatedTheta1 / (2+updatedTheta1)
    updatedY0 = np.random.binomial(n, p, 1)[0]
    thetas1.append(updatedTheta1)
    theta1 = updatedTheta1

print(np.mean(thetas1))
n, bins, patches = plt.hist(x=thetas1,density=True, bins='auto',
color='#0504aa',alpha=0.7, rwidth=0.85)
#Produce the 97.5 to generate the 95% equitable set
print(f"Our 95 percent equitable set is: {np.percentile(thetas1,
[2.5,97.5])}")
```

b) When running this code we can see that the following density plot and be derived for our posterior distribution:



c)

The equitable set is approximately:

```
Our 95 percent equitable set is: [0.38645041 0.5856323 ]
```