12.

In this question we will calculate the MAP estimate of the Bernoulli distribution using a non-conjugate prior. We will work with a prior that makes very strong assumptions about the behaviour of the coin (i.e. the coin is fair or slightly biased towards tails). Intuitively it makes sense that strong priors give us better results, if our initial believe is on the mark. However, strong priors can also impose a difficulty for the empirical data to shape the posterior.

Intro
Solution
a)
The expression for the new prior is given by:

$$p(\theta) = 0.5\delta(\theta - 0.5) + 0.5\delta(\theta - 0.4) \tag{1}$$

The likelihood of a given dataset remains the same, despite the changes in the prior:

$$p(D|\theta) = \theta^{N_1}(1 - \theta)^{N - N_1} \tag{2}$$

Therefore, the posterior has the following form:

$$
\begin{aligned}
p(\theta|D) &\propto p(D|\theta)p(\theta) = \theta^{N_1}(1 - \theta)^{N - N_1}(0.5\delta(\theta - 0.5) + 0.5\delta(\theta - 0.4)) = \\
&0.5^{N+1}\delta(\theta - 0.5) + 0.5(0.4)^{N_1}(0.6)^{N - N_1}\delta(\theta - 0.4) \\
&\propto 0.5^N\delta(\theta - 0.5) + (0.4)^{N_1}(0.6)^{N - N_1}\delta(\theta - 0.4)
\end{aligned} \tag{3}
$$

In (3), we substitute θ for the corresponding value of the indicator function (I). Note that the posterior only give us two non zero choices: $\theta = 0.5$ or $\theta = 0.4$. So, the MAP estimate will be the θ which gives us the biggest probability. Now we need to calculate under which conditions we should choose one or the other. Let's start assuming that $\theta_{MAP} = 0.5$:

$$
\begin{aligned}
&0.5^N > 0.4^{N_1}0.6^{N - N_1} \\
&N log(0.5) > N_1 log(0.4) + (N - N_1)log(0.6) \\
&N log(\frac{5}{6}) > N_1 log(\frac{2}{3}) \\
&N_1 > \frac{log(1.2)}{log(1.5)}N \approx 0.45N
\end{aligned} \tag{4}
$$

Since all the inequalities steps above are reversable we can state: $N_1 > 0.45N \iff \theta_{MAP} = 0.5$. A direct consequence of this statement is: $N_1 \leq 0.45N \iff \theta_{MAP} = 0.4$.

b)
The true parameter is $\theta = 0.41$. Let's analyse which prior give us the better estimate, if we are working with small datasets.

Small datasets
Old prior
Remember that for the old prior: $\theta_{MAP} = \frac{N_1 + \alpha - 1}{N + \alpha + \beta - 2}$ For small datasets, there is no guarantee that the ratio $\frac{N_1}{N}$ converges to the mean, which is equal

to the parameter θ in the Bernoulli distribution. Moreover, the influence of the prior hyper parameters is significantly higher when the dataset is small. Thus, the MAP estimate can end up being very bad.

New prior

With the new prior, the worst case scenario will be $\theta_{MAP} = 0.5$, which is not bad at all, considering we have a small dataset and the $\theta_{true} = 0.41$

Veredict

The new prior is better than the old one for small datasets

Now, let's make a similar analysis for large datasets.

Large datasets

Old prior

For large datasets, we can make the following approximation: $\theta_{MAP} = \frac{N_1+\alpha-1}{N+\alpha+\beta-2} \approx \frac{N_1}{N} \approx E[\theta] = \theta_{true}$. Therefore, the MAP estimate will be a pretty good estimation of the true parameter

New prior

No matter how large the dataset, our best guest will always be $\theta_{MAP} = 0.40$, given we only have two options to choose from.

Veredict

The old prior is better for large datasets.

Conclusion

In this question we calculated the MAP estimate for the Bernoulli distribution using a non-conjugate prior. We also compared the result of this MAP estimate with the result of the conjugate prior.